Volume 42 *September 1956* *Number 9*

Proceedings of the
National Academy of Sciences
of the United States of America

The Proceedings of the National Academy of Sciences

OFFICERS OF THE ACADEMY

DETLEV W. BRONK
President

GEORGE W. CORNER
Vice President

JOHN G. KIRKWOOD
Foreign Secretary

HUGH L. DRYDEN
Home Secretary

WILLIAM J. ROBBINS
Treasurer

EDITORIAL BOARD OF THE PROCEEDINGS

WENDELL M. STANLEY
Chairman

HUGH L. DRYDEN, *Home Secretary* JOHN G. KIRKWOOD, *Foreign Secretary*
EDWIN B. WILSON, *Managing Editor*

J. W. BEAMS	G. W. BEADLE	GREGORY BREIT
R. E. CLELAND	E. A. DOISY	L. C. DUNN
W. W. RUBEY	G. C. EVANS	W. F. LIBBY
HARLOW SHAPLEY	C. H. GRAHAM	J. VON NEUMANN
F. E. TERMAN	J. T. PATTERSON	PAUL WEISS

Subscription price is $12.50 per volume (12 issues); the price of a single issue is $1.25. Make all remittances payable to The University of Chicago Press in U.S. currency or its equivalent. Authorized agent for the British Commonwealth, except North America and Australasia: Cambridge University Press, Bentley House, 200 Euston Road, London, N. W. 1, England.

Editorial correspondence should be addressed to the PROCEEDINGS of the NATIONAL ACADEMY OF SCIENCES, Attention Miss Mary D. Alexander, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois.

Business correspondence should be addressed to the Journals Division, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois.

Subscribers are requested to notify The University of Chicago Press and their local postmaster immediately of change of address.

Microfilms of complete volumes of this journal are available to regular subscribers only and may be obtained at the end of the volume year from University Microfilms, 313 N. First Street, Ann Arbor, Michigan.

Entered as second-class matter June 21, 1920 at the Post-office at Easton, Pennsylvania, under the Act of August 24, 1912. Accepted for mailing at a special rate of postage provided for in Section 1103, Act of October 3, 1917. Authorized June 18, 1920.

PRINTED IN THE U. S. A.

THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES
is published for
THE NATIONAL ACADEMY OF SCIENCES
by
THE UNIVERSITY OF CHICAGO PRESS

Proceedings of the
NATIONAL ACADEMY OF SCIENCES

Volume 42 · Number 9 · September 15, 1956

*A NERVE GROWTH-STIMULATING FACTOR ISOLATED FROM SNAKE VENOM**

By Stanley Cohen and Rita Levi-Montalcini

DEPARTMENT OF ZOÖLOGY, WASHINGTON UNIVERSITY, ST. LOUIS, MISSOURI

Communicated by V. Hamburger, July 19, 1956

Introduction.—In the past our efforts have been directed toward the purification of an agent contained in mouse sarcomas 180 and 37 which has remarkable growth-promoting effects on sympathetic and spinal ganglia of the chick embryo.[1, 2, 3] In the course of attempts to characterize this material, use was made of crude snake venom as a source of phosphodiesterase. Treatment of the sarcoma factor with snake venom enhanced its activity, and subsequent tests showed that snake venom alone contains a very potent growth-promoting agent. In fact, it can be shown that crude venom is approximately 3,000–6,000 times as active as crude tumor homogenates (on a dry-weight basis) in promoting nerve fiber outgrowth in spinal ganglia in vitro. In the following, we are presenting the results of a partial purification and characterization of the snake-venom factor.

Materials and Methods.—Dried moccasin snake venom (*Agkistrodon piscivorus*) was obtained from the Ross Allen Reptile Institute. For the assay of the nerve growth-promoting activity, hanging-drop tissue cultures were made containing plasma (rooster), synthetic medium 1066 (with thrombin), and the material to be tested. The details of the procedure have been presented elsewhere.[3] Each culture contained three or four sensory ganglia isolated from 7-day chick embryos. The cultures were observed after 18 hours of incubation at 37°, and the growth of the fibers was semiquantitatively recorded from 1+ to 4+ (see Figs. 1–4 of our previous publication[2]). The assay was sensitive to twofold changes in concentration of the active material; smaller changes were not detectable.

Experimental Results.—The effect of crude venom when it is added to the cultures at a concentration of 7 μg/ml is shown in Figures 1 and 2. The active component of the venom was nondialyzable. Solutions of the venom (30 μg/ml in distilled water) were then treated with alkali (0.1 *N* sodium hydroxide for 1 hour at 26°), acid (0.1 *N* hydrochloric acid for 1 hour at 26°), 6 *N* urea (1 hour at 0°), and heat (5 minutes at 90°) in a manner identical to that described elsewhere[3] for the tumor factor. The biological activity was stable to urea and alkali and was completely destroyed by the heat and acid treatments. This behavior is similar to that shown by the "protein fraction" isolated from sarcoma 180.

In exploratory experiments it was found that the effect of the venom was abolished by the anti-snake-venom serum.[4] In these experiments a mixture of venom (20 μg/ml) and antiserum (40 μg/ml) was allowed to stand for 1 hour at 26°. Aliquots

were then tested for biological activity. The inhibitory effect of the antiserum suggests that the active factor is associated with one of the protein components of the venom.

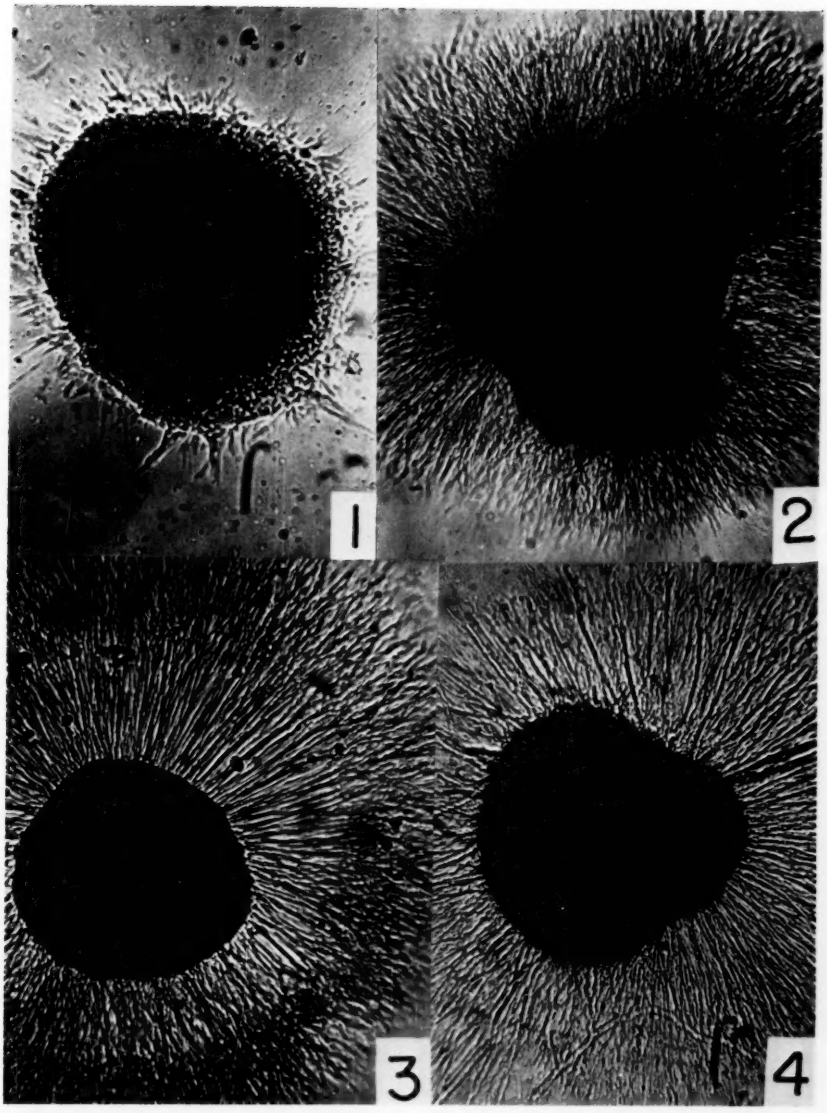
When the solution of the crude venom was allowed to stand overnight at 20°, approximately three-fourths of the activity was destroyed. However, if the crude venom was treated with 6 *N* urea, its stability was increased, and only slight loss of activity (less than 25 per cent) was noted after standing for 18 hours at 20°. This may be due to the destruction by the urea of an inactivating enzyme in the venom. Advantage of this was taken during the following purification procedure.

One gram of venom (*A. piscivorus*) and 36 gm. of urea were dissolved in 80 ml. of water containing 3 ml. of 0.1 *N* NaOH. The solution was allowed to stand for 90 minutes at 0°. All subsequent procedures were carried out at temperatures from 0° to 3°. Saturated ammonium sulfate (at 0°, pH 7.6) was added to the solution until a final concentration of 48 per cent saturation was obtained. The mixture was allowed to stand for 15 minutes and was then centrifuged for five minutes at 16,000 × *g*. The precipitate was discarded. To the supernatant fluid ammonium sulphate was added to a final concentration of 71 per cent saturation. The mixture was again allowed to stand for 15 minutes and then was centrifuged. The yellowish supernatant fluid was discarded. The precipitate, containing the active factor, was dissolved in 40 ml. of water and dialyzed against distilled water overnight. The material was then centrifuged for 10 minutes at 16,000 × *g*, and the slight precipitate was discarded. The supernatant fluid was again fractionated with the above saturated ammonium sulfate solution (in the absence of urea); the active material precipitated between 46 and 59 per cent saturation. After centrifugation the precipitate was dissolved in 20 ml. of water and dialyzed overnight against distilled water. This material may be stored in the frozen condition for at least several weeks with no detectable loss in activity. Ninety-five milligrams (dry weight) and approximately 40 per cent of the original total activity was recovered; a fourfold purification was thus achieved.

The biological activity of this material is more stable to the denaturing effects of the organic solvents than was the factor purified from sarcoma 180. To solutions containing 30 μg/ml were added ethanol, methanol, and acetone, to a final concentration of 66 per cent. The mixtures were allowed to stand for 3 hours at 30°. The organic solvents were removed by vacuum distillation at 3° and dialysis. Over 50 per cent of the original activity was retained. Similar treatment almost completely inactivated the tumor factor.[2]

The 280/260 mμ absorption ratio in distilled water, pH 7.0, was 1.75. The addition of streptomycin (0.01 *M*) caused no visible precipitation. These data indicate that only traces, if any, of nucleic acids were present. One milligram of this material was equivalent to 1.3 mg. of protein as determined by the procedure of Lowry *et al.*,[5] using bovine albumin as a standard; subsequent fractionations were based on protein content.

The venom factor was then further purified by adsorption and elution from the anionic and cationic cellulose resins DEAE-SF (1.1 meq/gm) and CM-cellulose (0.5 meq/gm).[6] Ten milliliters of the ammonium sulfate fraction (containing 3.3 mg. of protein per milliliter) were passed through a column 1.4 cm. in diameter containing 500 mg. of the DEAE-SF at a flow rate of 0.1–0.2 ml/minute. The column was then washed with 10 ml. of distilled water. Only traces of protein and



FIGS. 1–4.—Microphotographs of living sensory ganglia (from 7-day chick embryos) after 18 hours of incubation at 37°. Fig. 1, the standard control medium was used. Fig. 2, the medium contained 7 μg/ml of crude snake venom. Fig. 3, the medium contained 0.3 μg/ml of the CM fraction obtained from the venom. Fig. 4. The medium contained 250 μg/ml of the "protein fraction" purified from sarcoma 180.

no activity were present in this combined fraction. The active material was then eluted with 20 ml. of 0.01 *M* sodium chloride in 0.01 *M* potassium phosphate buffer, pH 7.4. This fraction (DEAE fraction) contained 15 mg. of protein and almost all (over 75 per cent) of the biological activity. The 18 mg. of protein which remained on the column could be quantitatively eluted with 20 ml. of 0.5 *M* sodium chloride in the phosphate buffer. This fraction was biologically inactive.

The active fraction was then dialyzed with stirring for 5 hours against distilled water, and 10 ml. containing a total of 7 mg. of protein were applied to a column 1.4 cm. in diameter containing 500 mg. of CM-cellulose at a flow rate of 0.1–0.2 ml/minute. The column was then washed with 10 ml. of water. This combined fraction contained 4.1 mg. of protein and was inactive. The activity was then eluted with 10 ml. of 0.1 *M* sodium chloride in 0.01 *M* potassium phosphate buffer, pH 7.4. This fraction (CM fraction) contained 1.9 mg. of protein and again almost all the activity (over 75 per cent). The remaining 1 mg. of protein could be eluted with 0.5 *M* sodium chloride in phosphate buffer and was inactive.

To summarize: approximately 25 per cent of the activity present in 1 gm. of the crude venom was recovered in 11.7 mg. of a "protein" fraction, a purification of approximately twenty fold.

The effect of the CM fraction on sensory ganglia in tissue cultures at a protein concentration of 0.3 μg/ml is shown in Figure 3. For comparison, the effect of the "protein fraction" obtained from sarcoma 180 at a concentration of 250 μg/ml is shown in Figure 4. The effect of the CM fraction in promoting the outgrowth of nerve fibers in the spinal ganglia appears to be identical with that of the tumor factor. In addition, as will be shown in another publication,[7] the injection of the DEAE fraction into the yolk of the chick embryo duplicates the nerve growth-promoting properties of the sarcomas in the embryo.

The venom from the two species examined, the moccasin (*A. piscivorus*) and the rattlesnake (*Crotalus adamanteus*), were equally effective in promoting fiber outgrowth in tissue culture.

The question arises concerning the identity of the factors purified from sarcoma 180 and venom. It would seem highly improbable that two chemically identical proteins would be found in such unrelated materials. The possibility that both factors have a common enzymatic activity is under investigation.

Summary.—A heat-labile, nondialyzable factor with very potent nerve growth-stimulating properties has been partially purified from snake venom. The biological effects of the venom factor, tested in tissue culture, are very similar to the effects produced by a factor obtained from sarcoma 180. However, our most purified venom preparations have a specific activity (on a protein basis) approximately 1,000 times as high as our purest tumor fraction. The data suggest that in both instances the active material is a protein or is bound to a protein.

* This work has been supported by Grant B-463 from the National Institute of Neurological Diseases and Blindness of the National Institutes of Health, Public Health Service, by a grant of the American Cancer Society, recommended by the Committee on Growth, National Research Council, and by a contribution from an institutional grant of the American Cancer Society to Washington University. The able assistance of Miss Barbara Bankhead is gratefully acknowledged.

[1] R. Levi-Montalcini, H. Meyer, and V. Hamburger, *Cancer Research*, **14**, 49–57, 1954.

[2] S. Cohen, R. Levi-Montalcini, and V. Hamburger, these PROCEEDINGS, **40**, 1014–1018, 1954.

[3] S. Cohen and R. Levi-Montalcini, *Cancer Research* (in press).

[4] Antivenin (Crotalidae) polyvalent was obtained from Wyeth Laboratories, Inc.

[5] O. H. Lowry, N. J. Rosebrough, A. L. Farr, R. J. Randall, *J. Biol. Chem.*, **193**, 265–275, 1951.

[6] We wish to thank Drs. A. Kornberg and H. Pahl for the diethylaminoethyl (DEAE) and carboxymethyl (CM) cellulose ion exchangers.

[7] R. Levi-Montalcini and S. Cohen, these PROCEEDINGS, **42**, 695–699, 1956.

---

*BACTERIAL PROTOPLASTS INDUCED BY PENICILLIN**

BY JOSHUA LEDERBERG

DEPARTMENT OF GENETICS, UNIVERSITY OF WISCONSIN, MADISON, WISCONSIN

Communicated by Sewall Wright, May 28, 1956

Bacterial protoplasts are believed to be cellular units that have been deprived of their rigid cell wall.[1] Accordingly, they are distinguished by their spherical shape (in bacilliform species) and their sensitivity to cytolysis in hypotonic media. Several authors[1–3] have suggested that protoplasts might be useful material for the study of biosynthesis of viruses and enzymes in a system simpler or more accessible to external modification than the intact cell. So far, only gram-positive bacteria, which are susceptible to lysozyme, have been used for such studies. A method has now been found for the efficient production of protoplasts from enteric bacteria such as *Escherichia coli* and *Salmonella typhimurium*, which are already familiar physiological and genetic subjects. The technique consists essentially of the exposure of growing cells to a medium containing penicillin, sucrose, and Mg^{++}.

Preparation of Protoplasts.—Escherichia coli strain K-12 and a variety of its mutant substrains are used in most of the experiments. The use of penicillin and sucrose was suggested by the possible analogy between penicillin and lysozyme as lytic agents and by the finding that hypertonic sucrose would interrupt bacteriolysis of *Bacillus megaterium* by lysozyme.[1] In addition, spherical bodies had been casually noted in other applications of penicillin,[4] and many authors have emphasized its use in the production of *L*-forms.[5] The following procedure was adopted after empirical trials and can doubtless be further improved.

The bacteria were grown overnight in tubes with 10 ml. of broth (Difco penassay medium) at 37° C., on a rotator. Samples of 3 ml. of the grown culture (about 2×10^9 cells/ml) were added directly to 10 ml. of broth supplemented with penicillin, 1,000 u/ml, sucrose 20 per cent, and magnesium sulfate 0.2 per cent. In 2–3 hours the cells were quantitatively converted into spheres. During this interval the optical density (measured at 650 mμ in a Coleman 14 spectrophotometer) increased about 50 per cent, but the total count (spheres or rods, estimated in a Petroff-Hausser chamber) remained constant. The spheres promptly lysed when the suspension was diluted in distilled water, and they are therefore regarded as "protoplasts."

The indicated supplements are in substantial excess, and nearly optimal yields of protoplasts can be obtained with 5 per cent sucrose, 100 u. penicillin, and 0.1 per cent $MgSO_4$. The high magnesium requirement may depend partly on binding with the sodium citrate used in the compounding of the penicillin preparation

(Squibb penicillin G, potassium, buffered). The Mg^{++} can be replaced by Ca^{++} but was preferred, to minimize precipitation of phosphate. In the absence of Mg^{++}, protoplasts are formed but are less well preserved, and a third or less of the initial rods are recovered. In the absence of sucrose or comparable stabilizer, the cells are almost totally lysed, leaving only debris. In the absence of penicillin, the rods grow at nearly the normal rate, i.e., are not appreciably inhibited by 20 per cent sucrose.

Protoplasts were also formed in minimal medium, but the transformation has been irregular and incomplete in the trials to date. Cells exposed to penicillin under conditions not supporting growth, e.g., in sucrose buffer, were not transformed into spheres.

Microscopic Observations and Viability.—The transformation of rods into spheres was observed in small droplets each containing ten to twenty cells, immersed in an oil chamber.[6] Each rod gave rise to a single sphere; even cells about to divide were inhibited in further growth and division. The progression of stages as observed with dark phase contrast at 645× is illustrated in Figure 1. A swelling first appears, either subterminally or (in incipient division stages) centrally, and progressively enlarges. The remainder of the rod then withers away and disappears.

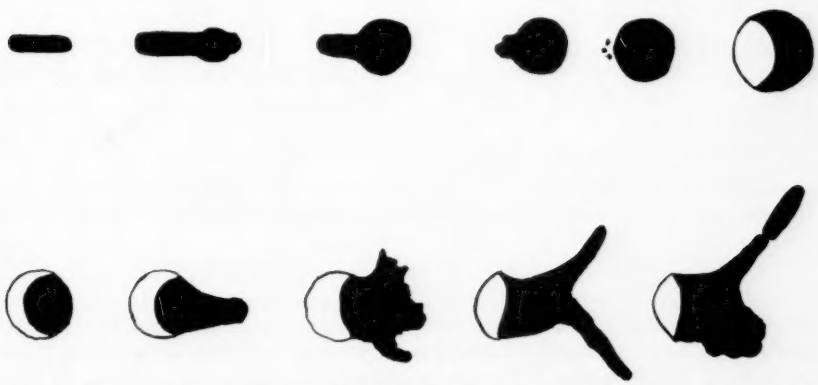


FIG. 1.—*Above:* Transformation of bacterial rod to spherical protoplast in the presence of penicillin. *Below:* reversion of protoplast to rod in absence of penicillin. Observed under dark phase contrast.

The reversion of spherical protoplasts to rods, in the absence of penicillin, has been followed by a similar technique. The spheres first enlarge, then develop amoeboid, finally rhizoid, outgrowths which segment to give typical rods. Spheres incubated in the penicillin medium continue to enlarge but vacuolate and eventually lyse over an interval of 24 hours. The optical density of the suspensions shows a parallel cycle. The enlarging spheres are highly reminiscent of published figures of *L*-forms,[5] but no evidence for multiplication of spherical elements has been seen so far in these experiments or in trials in which 10 per cent bovine or equine serum was furnished. The protoplast suspensions remain intact at least for several days at 4° C. After spontaneous lysis, or upon cytolysis in water, a residual ghost is seen

which retains its spherical form, although deprived of phase-dense material. The lysates are quite viscous, presumably from the release of desoxyribonucleic acid.

About half the protoplasts are capable of reverting to rods and engendering typical bacillary colonies in the absence of penicillin. This was verified by comparisons of total with viable counts on EMB lactose agar plus 20 per cent sucrose. The viability of the protoplasts is completely lost by cytolysis in water. On the other hand, normal rods are unimpaired when suspensions in sucrose solution are similarly diluted.

Enzyme Synthesis and Other Observations.—Protoplasts were tested for their ability to synthesize β-D-galactosidase: suspensions were sedimented, washed with 20 per cent sucrose, and resuspended in lactose medium (casein digest broth plus sucrose plus $MgSO_4$ plus penicillin plus 1 per cent lactose). Over an interval of several hours, samples were assayed with *o*-nitrophenyl galactoside.[7] In 210 minutes the enzyme activity increased from an initial value of 0.38 to 73 units per 10^9 spheres, a two-hundred-fold increase, while the optical density increased about twofold. (One unit corresponds to the splitting of 10^{-8} moles of substrate per minute. The peak activity found in spheres from lactose-grown cells is about 150 units. According to Rotman and Spiegelman,[8] each of these units corresponds to 5×10^{-9} gm. nitrogen, or $\sim 2 \times 10^{10}$ molecules of purified enzyme protein.) Most of the induced enzyme activity sedimented with the protoplasts but was released to the supernate when the protoplasts were cytolysed in water.

Protoplasts were produced in a similar fashion in *E. coli* strain B and in *S. typhimurium* strain TM2. The cells of the latter swam very actively, but the protoplasts were immotile, except for an occasional sphere on which a bacillary protuberance persisted (or had regenerated). Similarly, the protoplasts failed to agglutinate in homologous antiflagellar serum. However, electron-microscopic studies will be needed to establish the relationship of flagella to cell walls in this species. Protoplast suspensions gave a faint agglutination in homologous somatic antiserum.

The protoplasts of *E. coli* K-12 at first appeared to be readily stainable with methylene blue but were later found to be very sensitive to photodynamic effects which result either in lysis or stainability. Boiled protoplasts retain their form and are evidently fixed as they become much more opaque, stain readily, and do not cytolyse in water. Stained protoplasts often show a denser inclusion about one-half their diameter, a poorly staining cortex, and a deeper-staining periphery. The protoplasts reduce triphenyl-tetrazolium chloride, as well as methylene blue, very readily. With the former, a typical formazan granule is produced, usually single and near the periphery of the protoplast, much as in the rods.[9] No corresponding structure was seen in the untreated spheres.

In view of reports on the fusion of protoplasts,[10] and the possibility of DNA-mediated transduction, preparations have been made with a variety of appropriate genotypic mixtures of *E. coli* K-12. No evidence of hitherto unrecognized modes of genetic recombination has been uncovered in preliminary trials. Protoplast suspensions of appropriate compatibility types did, however, retain their ability to mate.

Discussion.—Most of the findings in this preliminary survey warrant more intensive and exact quantitative study. The areas of the interest already expressed

in bacterial protoplasts[1] are already beyond the scope of any one laboratory, and a ready method for their production in enteric bacteria may encourage further attention. In addition to their use in the extraction of intracellular constituents by cytolysis and their availability for biosynthetic studies (enzymes, viruses, cell walls), they may also be useful in cytological, antigenic, and physiological analysis. Their possible genetic applications are currently being studied in this laboratory.

The preservation of viable protoplasts in sucrose solution supports the hypothesis that the *primary* action of penicillin is on the synthesis or maintenance of a component of the cell wall,[11] while the activity and synthesis of other enzymes and over-all protoplasmic mass are unaffected.

The protoplasts of *E. coli* show some instructive differences from those of *B. megaterium*. The latter are rapidly produced by the enzymatic dissolution of the existing wall, and each rod produces not one but several spheres. The latter difference presumably corresponds to the septate structure of gram-positive rods, each compartment containing a single protoplast.[12] Gram-negative bacteria, on the other hand, are believed to be true multinucleate coenocytes. The most striking difference may vanish with technical improvements—namely, that the protoplasts of *B. megaterium* are reportedly inviable, while those of *E. coli* revert to normal rods, an obvious advantage for genetic research.

Summary.—Growing cells of *E. coli* or *S. typhimurium* are quantitatively transformed into spherical "protoplasts" in the presence of penicillin, magnesium, and high concentrations of sucrose. The protoplasts are quickly cytolysed in water but revert to viable rods in the absence of penicillin. The protoplasts have not been observed to divide as such but have been shown to enlarge and to synthesize an inducible enzyme (β-D-galactosidase) in appropriate media. The possible applications of these protoplasts in physiological and genetic research are cited.

* Paper No. 625 of the Department of Genetics. This work has been supported by grants from the National Science Foundation and the National Cancer Institute, Public Health Service (C-2157).

[1] E. T. C. Spooner and B. A. D. Stocker (eds.), *Bacterial Anatomy: Symposium of the Society for General Microbiology* (Cambridge: At the University Press, 1956).

[2] O. Landman and S. Spiegelman, these PROCEEDINGS, **41**, 698, 1955.

[3] S. Brenner and G. S. Stent, *Biochim. et biophys. acta*, **17**, 465, 1955.

[4] J. Lederberg and N. D. Zinder, *J. Am. Chem. Soc.*, **70**, 4267, 1948; J. Lederberg, E. M. Lederberg, N. D. Zinder, and E. R. Lively, *Cold Spring Harbor Symposia Quant. Biol.*, **16**, 413, 1953.

[5] L. Dienes and H. I. Weinberger, *Bacteriol. Revs.*, **15**, 245, 1951.

[6] J. Lederberg, *J. Bacteriol.*, **68**, 258, 1954.

[7] J. Lederberg, *ibid.*, **60**, 381, 1950.

[8] B. Rotman and S. Spiegelman, *J. Bacteriol.*, **68**, 419, 1954, and personal communication.

[9] J. Lederberg, *J. Bacteriol.*, **56**, 695, 1948.

[10] H. Stähelin, *Schweiz. Z. allgem. Pathol. u. Bakteriol.*, **17**, 296, 1954.

[11] P. D. Cooper, *Bacteriol. Revs.*, **20**, 28, 1956.

[12] C. F. Robinow, in R. J. Dubos (ed.), *The Bacterial Cell* (Cambridge, Mass.: Harvard University Press, 1945).

HINDERED CIS ISOMERS OF VITAMIN A AND RETINENE: THE STRUCTURE OF THE NEO-b ISOMER

By William Oroshnik, Paul K. Brown, Ruth Hubbard, and George Wald*

ORTHO RESEARCH FOUNDATION, RARITAN, NEW JERSEY, AND BIOLOGICAL LABORATORIES OF HARVARD UNIVERSITY, CAMBRIDGE, MASSACHUSETTS

Communicated July 9, 1956

Under the above title we have recently presented the following argument:[1] (*a*) The configurations of the four unhindered geometrical isomers of vitamin A and retinene having been established (Fig. 1),[2] the neo-b isomer proved to be a fifth form, necessarily therefore possessing a hindered *cis* linkage. (*b*) Isomerization studies showed it to be mono*cis*, therefore either 7- or 11-*cis*. (*c*) An 11-*cis* vitamin A (neo-c) having been synthesized and shown not to be neo-b, it was concluded that the latter must be 7-*cis*.

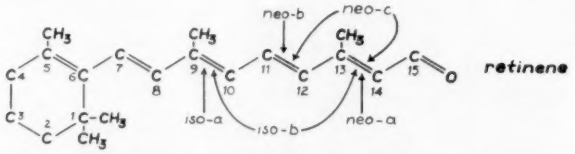


FIG. 1.—Structural formula of all-*trans* retinene. The arrows point to the positions of *cis* linkages in the other isomers, unhindered *cis* isomers below, hindered *cis* isomers above.

Compound II (Fig. 2), an intermediate in the commercial synthesis of vitamin A, was used to provide the terminal segment in our synthesis of neo-c vitamin A. This substance is prepared by the allylic rearrangement of compound I as a mixture of *cis* and *trans* isomers. We had used the predominant isomer, heretofore believed to be *trans*. On re-examination, however, this has now been shown to be *cis* and to have inadvertently introduced a second, terminal *cis* linkage into our synthesis. Neo-c vitamin A is therefore the 11,13-di*cis* isomer (Fig. 3).[3]

Robeson *et al.*[2] have proposed the 11,13-di*cis* structure for neo-b retinene. This suggestion seems to be confuted by our finding that this structure actually belongs to neo-c.

Using the *trans* isomer of compound II, the 11-mono*cis* isomer of vitamin A has now been synthesized (Fig. 3).[3] It is identical with neo-b vitamin A in ultraviolet and infrared spectra. It exhibits quantitatively the same relations in the antimony chloride reaction and on isomerization with light in the presence of iodine. A sample of this substance, oxidized to the corresponding aldehyde with manganese dioxide, displays quantitatively the same properties as neo-b retinene in the antimony chloride test and on isomerization with iodine and light.[4] Crystallized from petroleum ether, it melts at 63.1°–64.4° C. (neo-b retinene, 63.5°–64.4°;[2] 60.8°–62.3°[4]). The mixed melting point of this product with neo-b retinene was 61.5°–63.4°. A mixture of synthetic 11-*cis* retinene and cattle opsin, incubated in the dark, yields rhodopsin in the expected amount.[4]

It was noted earlier that neo-c retinene is sterically unstable, isomerizing slowly even at −15° C. in the dark.[1] The realization that this is the 11,13-di*cis* isomer

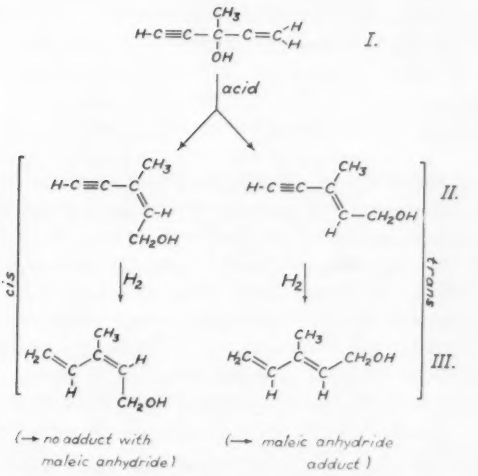
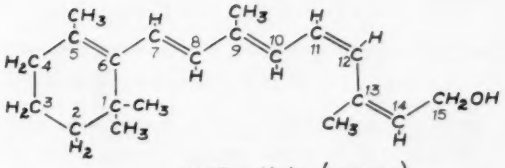


FIG. 2.—Allylic rearrangement of compound I on treatment with acid yields *cis* and *trans* isomers of compound II. After these have been reduced to the corresponding dienols (compound III), their configurations are established by testing with maleic anhydride. Only the *trans* isomer of compound III readily forms an adduct with this reagent.



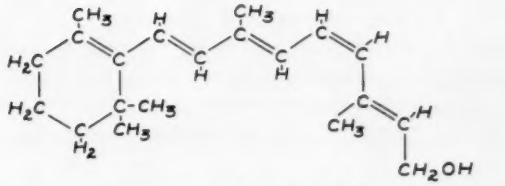


FIG. 3.—Structural formulas of neo-c and neo-b vitamin A. The 11-*cis* linkage is hindered in both molecules, owing to overlap between the 10-H and 13-CH_3 groups.

clarifies what was previously a puzzling observation. A preparation of neo-c retinene, partially purified chromatographically, was concentrated in petroleum ether and stored at −15° C. It partly crystallized; but on examination the crystals turned out to be neo-a (13-*cis*). It is now clear that during storage the relatively unstable hindered 11-*cis* linkage had largely isomerized to *trans*, yielding this result.

The positions of the long-wave-length absorption maxima (λ_{max}) of the neo-b and neo-c isomers are unusual. A hindered *cis* linkage in conjugated polyenes normally shifts λ_{max} markedly toward shorter wave lengths. In neo-b vitamin A and retinene, however, the displacement is only 5–7 mμ, the same as with an unhindered *cis* linkage. A terminal *cis* linkage, as in 13-*cis* (neo-a) or 9,13-di*cis* (iso-b) vitamin A, has heretofore been found to shift λ_{max} 1–3 *m*μ toward *longer* wave lengths, as compared with terminal *trans* configurations. The λ_{max} of neo-c vitamin A might be expected therefore to be 1–3 mμ longer than that of neo-b; instead, it is 8–10 mμ shorter. The reason for these unexpected relationships is not immediately apparent.

* This investigation was supported in part by grants to George Wald from the Rockefeller Foundation and the Office of Naval Research.

[1] G. Wald, P. K. Brown, R. Hubbard, and W. Oroshnik, these PROCEEDINGS, **41**, 438, 1955.

[2] C. D. Robeson, J. D. Cawley, L. Weisler, M. H. Stern, C. C. Eddinger, and A. J. Chechak, *J. Am. Chem. Soc.*, **77**, 4111, 1955; C. D. Robeson, W. P. Blum, J. M. Dieterle, J. D. Cawley and J. G. Baxter, *J. Am. Chem. Soc.*, **77**, 4120, 1955.

[3] W. Oroshnik, *J. Am. Chem. Soc.*, **78**, 2651, 1956.

[4] P. K. Brown and G. Wald, *J. Biol. Chem.* (in press).

---

*ON THE SYNTHESIS OF ANTIBODY PROTEIN IN VITRO**

BY DONALD F. STEINER† AND H. S. ANKER

DEPARTMENT OF BIOCHEMISTRY, UNIVERSITY OF CHICAGO

Communicated by W. H. Taliaferro, July 10, 1956

It is generally agreed that antibody formation is associated with organs containing relatively large numbers of lymphoid cells.[1] Among these organs, the spleen, in particular, participates in antibody formation, especially when antigens are administered intravenously. Formation of antibody also occurs in nonimmunized, immature, or X-irradiated recipient animals into which spleen cells from immunized donors are injected. Previous observations[2] indicate that in the rabbit the major proportion of antibody protein in an anamnestic response is synthesized between the third and fifth days following intravenous injection of a soluble antigen.

In vitro observations with spleen explants or slices have been restricted thus far chiefly to hemolysins,[3, 4] typhoid[5] and paratyphoid[6] agglutinins, and diphtheria antitoxins.[7]

A method is reported here which permits the synthesis of antibody proteins by rabbit spleen cells on incubation for several days in vitro. In these preparations the cells are maintained with a minimum of degenerative changes and form precipitable antibody against a soluble antigen in quantities comparable to the amounts found in in vivo transfer experiments.

MATERIALS AND METHODS

Immunization Procedure.—Young adult male albino rabbits weighing 2–3 kg. were fed ad libitum. The rabbits received a single primary injection of 100 mg. of crystalline bovine serum albumin (BSA, Armour Laboratories) in saline. On the twelfth day blood was drawn from the ear vein and the serum titrated for antibody. A wide range of serum antibody levels was noted, but only anergic animals were excluded. After a period of at least 1 month after the primary injection, a second dose of approximately 15 mg. of BSA was administered, and after the desired interval the spleen was removed aseptically under nembutal anesthesia.

Analytic Procedures.—Antibody was determined by precipitation with antigen.[8] Quantitative protein estimations on the precipitates were made by the method of Lowry,[9] using standardized bovine serum albumin as a reference. Glucose was determined colorimetrically.[10]

Air-dried smears of all cell preparations were stained with Wright's stain, while wet preparations of living cells were observed directly under the microscope. In some experiments cell counts were made with a hemocytometer.

Incubation Methods.—The spleen was weighed and minced and the fragments pushed through a No. 80 stainless-steel mesh screen with a glass pestle into a measured volume of medium under sterile conditions. Single-cell suspensions with only a few cell clumps were thus obtained.

Eagle's tissue culture fluid[11] with increased salt concentration and 0.34 per cent glucose, Trowell's medium[12] with penicillin and streptomycin, and Medium No. 199 of Morgan *et al.*[13] were used. Varying amounts of serum were added to the media.

Procedure A.—A suspension of cells was gently shaken at 37° C. in 30 ml. of medium in 95 per cent O_2–5 per cent CO_2. At the end of the incubation period the flask contents were centrifuged and the cells extracted after disintegration by sonic vibration. The extract and the supernatant medium were analyzed for antibody.

Procedure B.—Figure 1 illustrates the incubation flask used in these experiments. Cellophane membranes were heated to boiling, soaked for 24 hours in distilled water, and stretched across the glass flange of the upper section. The flange of the lower section was greased with vaseline. After autoclaving, the flanges were fastened together by steel springs. The lower (reservoir) compartment was completely filled with sterile medium, so that the membrane was flat and horizontal. Five milliliters of cell suspension were poured into the upper compartment and evenly distributed on top of the membrane by swirling. The cells thus settled uniformly on the membrane. The flasks were incubated at 37° C. with magnetic stirring of the reservoir at 60 rpm; 95 per cent O_2–5 per cent CO_2 was continuously passed over the cell suspension. About 10 minutes elapsed between splenectomy and the beginning of incubation. Five milliliters of the same suspension served as a zero time control. After incubation, the fluid above the membrane was decanted and the cells adhering to the membrane examined histologically.

RESULTS

After very short periods cells incubated by procedure A began to show degenerative changes. The large immature lymphoid and reticular cells disintegrated, and only a few small lymphocytes were motile after 48 hours. Hypertonic media

did not prevent these degenerative changes. Glucose utilization under these conditions was much lower than in experiments where the cells remained at rest. No antibody was found either in the incubation medium or in the extract from disrupted cells. The results are tabulated in Experiments 1 and 2 in Table 1.

TABLE 1

Expt.	Day of Splenectomy	Period of Incubation (Hours)*	Medium†	Glucose Uptake (Mg/Gm Spleen/24 hrs.)	Antibody (Mg/ml) Medium	Antibody Calc. Mg/Gm Spleen
1‡	3	24	EM–5%RS	8	..	..
		48	EM–10%RS	..	..	..
		72	.....	..	..	..
2‡	3	48	EM–50%RS	..	..	..
2§	3	48	EM–50%RS	..	..	1.9
3§	3	11	EM–50%RS	..	0.29	..
		28	EM–50%RS	..	1.15	..
		49	EM–50%RS	..	1.46	15.0
4§	3	48	EM–50%RS	..	0.25	2.0
		72	EM–50%RS	..	0.25	2.0
5§	3	48	EM–50%RS	66	0.30	2.3
6§	2	72	EM–50%RS	25	0.15	0.65
7§	2	72	EM–50%RS	45	0.065	0.7
8§	2	72	EM–10%RS	..	0.1	..
			EM–50%HS	..	0.19	2.7
9§	3	48	EM–50%RS	..	1.0	6.0
			EM–50%RS	..	1.0	..
			EM–50%HS	..	0.18	..
10§	3	48	TM–50%HS	..	0.18	1.8
			TM only	..	0.14	..
11§‖	3	48	EM–50%RS	..	0.4	6.2

* Cell counts ranged from 10^7 to 10^8 per milliliter.
† EM, modified Eagle's medium; TM, Trowell's medium; RS, rabbit serum or HS, hog serum in reservoir.
‡ Incubation procedure A.
§ Incubation procedure B.
‖ 6 mg. *l*-C^{14}-glycine added to incubation medium; 250 counts/min/mg of dry BSA-antibody precipitate.

Incubation by procedure B preserved all cell types normally seen in the spleen (Fig. 2). In particular, reticular cells, large lymphoid cells, and myeloid cells with granules were present (Fig. 3) for as long as 72 hours. Large numbers of motile lymphocytes were seen in wet preparations. However, after incubation for 5 days without change of medium, many large mononuclear cells (polyblasts) and large multinucleated giant cells appeared (Fig. 4). No formation of fibroblasts occurred, and only few mitoses were seen.

Antibody was formed when cells removed *three* days following the secondary challenge were incubated for 48 hours (Expt. 2). From 0.18 to 1.46 mg. of antibody per milliliter of medium was found (i.e., 1.8–15 mg. per gram of spleen). No antibody could be detected either in the supernatant from the controls or in the extracts of the cells either before or after incubation. The antibody concentration in the medium continued to increase for 48 hours (Expt. 3) but remained constant after that (Expt. 4). Glucose consumption was higher than in procedure A (Expt. 5). Less antibody was found in experiments in which the spleen was removed *two* days following administration of the antigen (Expts. 6, 7, and 8), and no antibody was found either if the spleen was removed on the *first* day or if antigen was added in vitro.

Maximal formation of antibody was found when modified Eagle's or Trowell's medium plus 50 per cent serum was used. Rabbit serum was added to the upper

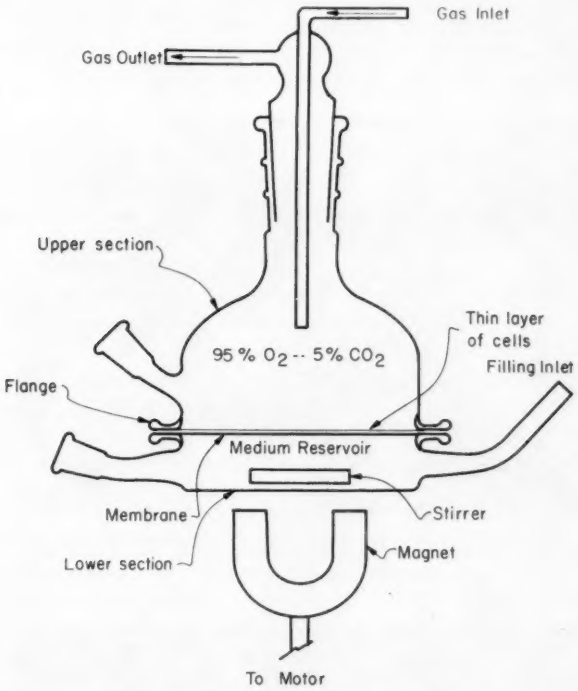


FIG. 1.—Incubation flask used in procedure B

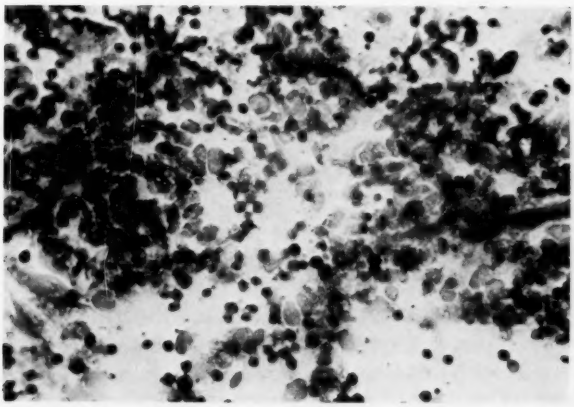


FIG. 2.—Thick air-dried smear of rabbit spleen cell suspension before incubation. Numerous reticulum cells and small cells of the lymphoid series. Wright stain, ×450.

section, and rabbit, human, or hog serum (Expt. 9) to the reservoir. Modified Eagle's medium alone was insufficient. In Trowell's medium, which contains higher concentrations of all the natural amino acids, approximately 75 per cent of

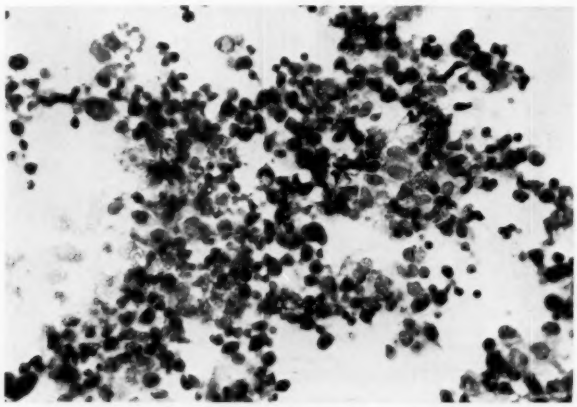


FIG. 3.—Thick air-dried smear from the membrane surface after 48 hours' incubation without change of medium. Numerous large reticulum cells; abundant smaller lymphoid cells. Wright stain, ×450.

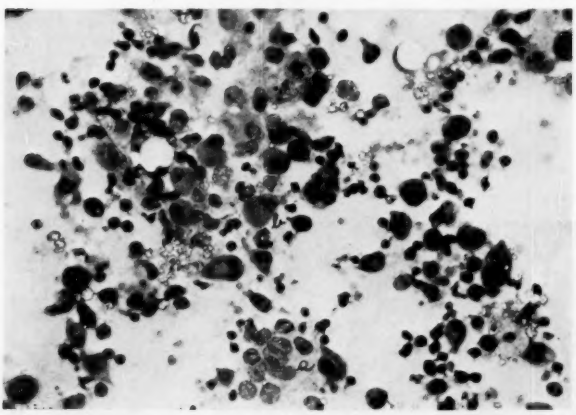


FIG. 4.—Thick air-dried smear from the membrane surface after 6 days of incubation without change of medium. Multinucleated giant cell (*a*) and numerous large polyblasts (*b*). Wright stain, ×450.

the maximal antibody production was observed, since the cells apparently survived for a sufficiently long period (Expt. 10). However, gross degenerative changes were evident histologically after 30 hours in all media in the absence of serum.

In Experiment 11, 6 mg. of *l*-C^{14}-glycine (3 μC/m*M*) was added. After incubation, equivalent amounts of egg albumin and anti–egg albumin antibody were added, the precipitate removed, and the anti-BSA antibody isolated by precipita-

tion with BSA. Both precipitates were dialyzed against a solution of unlabeled glycine. No C^{14} was found in the egg albumin precipitate. The isotope concentration of the BSA precipitate was about one-half that calculated assuming that all the glycine residues of the antibody were derived from the added labeled glycine.

DISCUSSION

It was found that incubation of surviving tissues by shaking in a suitable gas phase is unsatisfactory for prolonged survival of spleen cell suspensions, apparently because of mechanical damage of the cells by the agitation. Since antibody was not found, it appears that intact cells are required for the synthesis.

To avoid damage by agitation, spleen cells were incubated at rest. To overcome the limitations of gas diffusion in a large volume of medium, the cells were supported in a thin layer by a semipermeable membrane. The layer of fluid above the cells was only about 1 mm. thick. The large volume of medium in contact with the underside of the membrane served as a reservoir of nutrient materials and permitted the removal of metabolic products. Exchange of materials was enhanced by stirring. On incubation of spleen cell suspensions by this method, cell degeneration was minimized and antibody was formed in such amounts that it could be precipitated directly. Since the antibody protein could not diffuse through the membrane into the reservoir, it was found in all experiments only in the medium in which the cells were suspended.

Antibody formation was considerably lower if the spleen was removed 2 days after a secondary antigen injection. When the spleen was removed after 24 hours, no antibody formation could be detected, although all the cells were histologically fairly well preserved for 4 days. The formation of polyblasts and multinucleated giant cells after 5 days of incubation in vitro indicates that for this period of time the medium presently used does not sustain the cells in the same histological state as is maintained in the intact organ in vivo. The inadequacy of the culture medium probably accounts in part for lack of antibody formation when the spleen is removed prior to the second day after antigen administration. Similar effects had been reported previously[3] and attributed to a deficiency in the culture system.

Even though serum is not absolutely essential for the synthesis of antibody protein, extracellular protein appears to contribute to the preservation of the cells. It has likewise been noted that serum is necessary for maintenance of growth in many cell cultures.[11] It cannot be decided at the present time whether the serum effect is due to the protein itself or to low-molecular-weight factors associated with it.

Taliaferro and Talmadge[14] have studied the anamnestic response of rabbit spleen cells removed 3 days following a secondary challenge with BSA after transfer to nonimmunized recipients. In these experiments a total of from 4.5 to 50 mg. of antibody protein per spleen was synthesized following transfer of one spleen to one or two animals after a challenging dose of 60–90 mg. of antigen, or about 4–5 times the amounts used here. Although the magnitude of the antibody response as a function of the antigenic stimulus is not known, the amount of antibody protein synthesized in vitro in the experiments reported here ranged from 3 to 37 mg. per spleen. The yield in previous in vitro experiments[7] was only about one-tenth of the amount of antibody obtained in equivalent transfer experiments in vivo.

In all previous in vitro experiments[3-7] antibody formation could be demonstrated only because of the high potency of the antigens used and because of the very sensitive immunological methods available for the detection of antibodies against particulate antigens and toxins.

The method of incubation described here would seem to hold promise for other experiments requiring the maintenance of differentiated cells in a functional state in vitro.

SUMMARY

A technique has been developed which permits incubation at rest of large quantities of rabbit spleen cells in vitro. Under these conditions net synthesis of precipitating antibody against bovine serum albumin occurs. The amount of antibody formed corresponds closely to the amount obtained after transfer of spleen cells to nonimmunized recipients in vivo. It seems that the success of this technique is due to the prevention of damage by agitation and to provisions for adequate nutrition and aerobic metabolism of the cells.

* Supported by a grant from the United States Atomic Energy Commission.

† In partial fulfilment of the requirements for the M.S. degree.

[1] T. N. Harris and S. Harris, *Am. J. Med.*, **20,** 114, 1956.
[2] H. Green and H. S. Anker, *Biochim. et biophys. acta*, **13,** 365, 1954.
[3] R. C. Parker, *Methods of Tissue Culture*, 2nd ed.; New York: P. B. Hoeber, 1950).
[4] S. Roberts *et al.*, *J. Immunol.*, **62,** 155, 1949.
[5] A. Fagraeus, *J. Immunol.*, **58,** 1, 1948.
[6] G. J. Thorbecke and F. J. Keuning, *J. Immunol.*, **70,** 129, 1953.
[7] A. B. Stavitsky, *J. Immunol.*, **75,** 214, 1955.
[8] M. Heidelberger and C. F. C. MacPherson, *Science*, **97,** 405, 1953.
[9] O. R. Lowry *et al.*, *J. Biol. Chem.*, **193,** 265, 1951.
[10] N. Nelson, *J. Biol. Chem.*, **153,** 375, 1944.
[11] H. Eagle, *Science* ,**122,** 501, 1955.
[12] O. A. Trowell, *Exptl. Cell Research*, **9,** 258, 1955.
[13] J. F. Morgan, H. J. Morton, and R. C. Parker, *Proc. Soc. Exptl. Biol. Med.*, **73,** 1, 1950.
[14] W. H. Taliaferro and D. W. Talmadge, *J. Infectious Diseases*, **97,** 88, 1955.

PRODUCTION OF PROTOPLASTS OF ESCHERICHIA COLI BY LYSOZYME TREATMENT

BY NORTON D. ZINDER AND WILLIAM F. ARNDT

ROCKEFELLER INSTITUTE FOR MEDICAL RESEARCH

Communicated by René J. Dubos, July 6, 1956

Recently, investigators of some biochemical problems have made use of subcellular elements of bacteria in an attempt to study systems more highly integrated than soluble extracts but more amenable to external manipulation than intact cells. Cells which either have been sonically disrupted[1] or have had their cell walls removed (protoplasts)[2] have been used. Protoplasts have hitherto been prepared only from gram-positive organisms. Treatment of *Bacillus megaterium* with lysozyme in hypertonic media brings about a change from its normal rodlike form

to a spherical one and results in its susceptibility to lysis upon suspension in hypotonic media.[3] These altered properties are presumed to result from dissolution of the cell wall, the cytoplasmic contents then being retained by the cell membrane. Unfortunately, little is known of the genetics of the gram-positive organisms, and the fruitful concomitant biochemical and genetic attack is thus restricted. For this reason it would be of obvious advantage to obtain protoplasts of such organisms as *Escherichia coli*, which has been the subject of detailed genetic and physiological investigation. This report details the development of such a technique which also involves the use of lysozyme.

Our thinking on the problem of obtaining protoplasts of *E. coli* was influenced by the report of Hirsch[4] that some material isolated by him from rabbit leukocytes (phagocytin) would kill *E. coli* strains. Phagocytin had chemical properties similar to, but not identical with, those of lysozyme. When *E. coli* cells were treated with crude phagocytin, (kindly provided by Dr. Hirsch) in hypertonic media at pH 5, they became susceptible to lysis upon dilution into distilled water. However, only a rare cell was morphologically identifiable as a protoplast. The difficulty in obtaining phagocytin in sufficient quantity for detailed investigation caused us to seek another source of active material. On the hypothesis that phagocytin was a lysozyme of different specificity, hen egg albumen, a rich source of lysozyme, was tested. Treatment of cells with a one-to-four dilution of fresh albumen resulted in the rapid conversion of cells to protoplasts at pH 5 and 9. At pH 5 the spheres were badly agglutinated. All further experiments, unless otherwise noted, were therefore done at pH 9. No activity of albumen was detected at pH 6–8. Crude fractionation of egg albumen revealed that the activity was associated with the lysozyme fraction. Commercial lysozyme (Armour Chemical Company) was equally effective.

CONVERSION OF CELLS TO PROTOPLASTS

The following procedures for the efficient conversion of cells of *E. coli* strains K12 and B, and their derivatives, were empirically developed. Overnight broth cultures of cells are diluted tenfold in Difco Penassay broth and incubated with aeration for 90 minutes. This takes the culture just out of logarithmic growth phase (cells less fragile) and gives a titer of $4–5 \times 10^8$ cells per milliliter. The culture is washed once in saline by centrifugation and resuspended in saline. After 3 minutes of incubation at 37° C., sufficient tris buffer (Sigma Chemical Company) and lysozyme are added to make the final concentrations 1 *M* and 200 μg/ml., respectively. Further incubation for 5 minutes converts the cells to protoplasts (Fig. 1). Other supporting media that can be used are sucrose 0.6 *M* with 0.1 *M* carbonate buffer, pH 9, or *M*/2 NaCl with carbonate buffer, or 0.3 *M* NH_4Cl brought to pH 9 with NaOH. In these latter two media the cells will cytolyze upon dilution into hypotonic media but do not assume a spherical shape. Centrifugation of the preprotoplasts and resuspension in 0.6 *M* sucrose will cause their immediate swelling into spheres.

With regard to these procedures, some points need be mentioned: (1) At pH 8–8.8, protoplasts are formed but are badly agglutinated. This agglutination can be prevented by the addition of 0.1 per cent Mg^{++}; however, some 10 per cent of the cells retain their rodlike shape and are not subject to cytolysis. (2) It is requi-

site that the cells be washed and starved slightly for quantitative conversion to occur. Perhaps they can heal the lysozyme-induced lesions if they are able to grow. (3) Certain strains will not form spheres in the treatment medium but will do so when resuspended in other media. In order for the cells to form spheres, they apparently must imbibe some water and swell—in a sense, hatch out of their cell walls. As yet there is no rationale for the preferential use of any medium, other than its empirical success with the strain in question.

PROPERTIES OF PROTOPLASTS OF *E. coli*

1. *General.*—Growing cells of *E. coli* are generally in a doublet condition. Each doublet gives rise to two protoplasts, with separation occurring at the point of incipient division. Motile strains lose their motility when converted to protoplasts. The spheres do not detectably adsorb bacteriophage. The protoplasts readily reduce methylene blue.

When protoplasts are suspended in broth media, they lyse no matter how high the osmotic pressure of the medium and whether it is maintained by sucrose, NaCl, or dextran. Addition, to a final concentration of 2 per cent, of bovine serum albumin prevents their lysis and makes further study possible. The following medium (P-broth) was used to study the growth of bacteriophage by, and the viability of, the protoplasts: Difco Penassay broth was supplemented with 0.6 *M* sucrose, 2 per cent serum albumin, and 0.1 per cent Mg^{++} (this prevents agglutination at the lower pH of this medium and may be required for growth processes).

2. *Growth of Bacteriophage.*—Protoplasts of *E. coli* B will grow bacteriophage T4. Cells were grown as described, and suspended, after washing, in saline containing 100 μg/ml. *l*-tryptophan (necessary for phage adsorption). Phage was added so as to give one particle per hundred bacteria, and after 5 minutes of incubation for adsorption and injection the cells were converted to protoplasts in tris buffer. Infective center titers (plaques) were obtained for protoplasts, lysed protoplasts, and cells by plating in $M/2$ NaCl phage agar (8 per cent tryptone, 0.1 per cent glucose, and agar) with an excess of *E. coli* B. Suitable samples were incubated in P-broth to determine final yields. Table 1 shows the results of such an experiment. About 20 per cent of protoplast infective centers are recovered directly. Only 1 per cent of infective centers are recovered from lysed protoplasts. These probably represent unadsorbed and desorbed phage, as there is little rise in the titer with time. Protoplasts give almost as large a yield as the untreated cells, indicating that the initial loss of protoplast infective centers is due to their premature lysis on the plates rather than to inability to grow phage.

TABLE 1

GROWTH OF BACTERIOPHAGE T4 IN PROTOPLASTS AND CELLS OF *E. coli* B

(Number of Plaques per Plate, Had Equivalent Volumes of Equivalent Dilutions Been Plated)

	Incubation Time			Incubation Time	
	0 Min.	90 Min.		0 Min.	90 Min.
Cells	350	21,100	Lysed protoplasts	3	12
Protoplasts	83	15,500			

3. *Viability.*—For 4 hours after suspension and incubation in P-broth, protoplasts exhibit no gross morphological changes as observed by phase-contrast microscopy. At this time, the majority of the protoplasts start to shrink, become highly refractile at their periphery, and assume a coccobacillus form. Their fate is still uncertain, but they probably represent a degenerate state. On the other hand, about 10 per cent of the protoplasts increase in volume and form a large vacuole. From the side distal to the vacuole, finger-like processes appear which grow into rods (Fig. 2). The mother protoplast becomes extremely distorted,

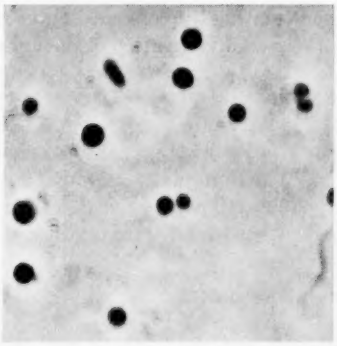


FIG. 1.—Phase-contrast microphotograph of wet mount of protoplasts. Note nonswollen rod for size comparison between the rods and spheres. ×1,700.

FIG. 2.—Phase-contrast microphotograph of wet mount of protoplast budding off a rod. ×1,700.

and a wide variety of morphological abberations are observed, reminiscent of the growth of L-forms from large bodies.[5] Direct demonstration of the viability of the protoplasts has been obtained by plating them in P-broth to which sufficient agar to just solidify the mixture was added. The equivalent of from 2 to 20 per cent (different experiments) of the treated cells formed normal colonies. On no other solid media have the protoplasts returned to the actual bacillary form.

DISCUSSION

Although lysozyme has been claimed to be without effect on *E. coli* cells, it was known to be bound by them.[6] In our studies it was clear that at pH 6–8 lysozyme was without effect, whereas its activity was extremely rapid at pH 8 and above. It is interesting to note that suspension of *E. coli* in the alkaline treatment medium, in the absence of lysozyme, brings about, within 30 minutes, the slow lysis of the cells upon their suspension in hypotonic media. This action may be analogous to, though less efficient than, that of lysozyme, which is itself a basic protein. Certain classes of bacteriolytic agents may have as their primary site of activity the cell-wall substance, with the lysis of the cell being due to secondary osmotic effects following the dissolution of the cell wall. In the course of these investigations it was noted that the lysis of *E. coli* by phage ghosts (phage membranes obtained by osmotic shock) went through a protoplast stage. However, no suitable stabilizing medium for such protoplasts was found. Thus three biological reagents—phagocytin,

lysozyme, and phage ghosts—are available for efficient lysis of *E. coli* cells for preparation of bacterial enzymes, nucleic acids, cell membranes, etc.

The protoplasts of *E. coli* have retained—as had their predecessors, the protoplasts of *B. megaterium*[7]—sufficient of their synthetic capacities to synthesize bacteriophage. This may be taken as a general indication of their biochemical integrity. The difference in the viability of *E. coli* and *B. megaterium* protoplasts may be only a reflection of the different media used in the different studies. As pointed out, there has been considerable variation in the viability of the *E. coli* protoplasts. This may depend on the amount of residual cell-wall substance left on the protoplast and/or the soluble materials lost during the conversion process.

With the ready accessibility of protoplasts of *E. coli*, investigation of their genetic potentialities is under way.

SUMMARY

The cells of two strains of *E. coli* can be quantitatively converted into spherical protoplasts by the action of lysozyme at pH 9. The protoplasts are capable of all the biosyntheses necessary for phage growth. About 10 per cent of them are also capable of reverting to normal rod forms when incubated in an appropriate medium.

[1] E. F. Gale and J. Folkes, *Nature*, **173**, 1223, 1954.
[2] O. Landman and S. Spiegelman, these PROCEEDINGS, **41**, 698, 1955.
[3] C. Weibull, *J. Bacteriol.*, **66**, 688, 1953.
[4] J. G. Hirsch, *J. Exptl. Med.*, **103**, 598, 1956.
[5] L. Dienes and H. I. Weinberger, *Bacteriol. Revs.*, **15**, 245, 1951.
[6] E. H. Boasson, *J. Immunol.*, **34**, 281, 1938.
[7] S. Brenner and G. S. Stent, *Biochim. et biophys. acta*, **17**, 473, 1955; M. R. J. Salton and K. M. McQuillen, *Biochim. et biophys. acta*, **17**, 465, 1955.

ION ASSOCIATION IN AQUEOUS SOLUTIONS OF 1-1 ELECTROLYTES AT 25°

BY MARIO AZZARRI* AND CHARLES A. KRAUS

METCALF RESEARCH LABORATORY, BROWN UNIVERSITY

Communicated June 22, 1956

According to Onsager's theory, the conductance of an unassociated electrolyte in aqueous solution decreases with concentration in accordance with the equation

$$\Lambda = \Lambda_0 - (\alpha\Lambda_0 + \beta)\sqrt{C}. \quad (1)$$

It has been found, however, that, while the conductance approaches the theoretical value in this solvent as C approaches zero, the observed values of Λ at higher concentrations are somewhat higher than those predicted by theory. The deviations from theoretical values increase with concentration approximately as a linear function of C. We thus have Shedlovsky's empirical extension[1] of Onsager's equation.

$$\Lambda = \Lambda_0 - (\Lambda_0\alpha + \beta)\sqrt{C} + BC, \quad (2)$$

where B is an empirical constant.

While conductance values of inorganic salts quite generally conform to equation (2) with a positive value of B, this is not the case with many quaternary ammonium salts.[2] It has been suggested that the observed negative deviations from Onsager's equation are due to ion association. However, up to now, the correctness of this hypothesis has not been established. It is our purpose, in this paper, to determine whether or not ion association actually occurs in aqueous solutions.

Before proceeding to an examination of available data, we need to consider the equilibrium in a dilute electrolyte solution in which ion association occurs to a very small extent.

If ion association occurs, the conductance is given by the equation

$$\Lambda = \Lambda_0 - (\alpha\Lambda_0 + \beta)\sqrt{C} + BC + \Delta_a\Lambda, \tag{3}$$

where $\Delta_a\Lambda$ is the conductance decrease due to ion association; thus it is always negative. If $\Delta_0\Lambda$ is the deviation from the simple Onsager equation, $\Delta_a\Lambda$ is given by the equation

$$\Delta_a\Lambda = \Delta_0\Lambda - BC. \tag{4}$$

The value of $\Delta_0\Lambda$ is approximated fairly closely by the equation

$$\Delta_0\Lambda = \Lambda'_0 - \Lambda_0, \tag{5}$$

where

$$A'_0 \equiv \frac{A + \beta\sqrt{C}}{1 - \alpha\sqrt{C}}. \tag{6}$$

Up to 0.01 N, the values of $\Delta_0\Lambda$ may be quite closely approximated by means of equations (5) and (6). $\Delta_0\Lambda$ is usually, but not always, negative if ion association occurs.

The degree of association, $1 - \gamma$, is given by the equation

$$1 = \gamma = \frac{-\Delta_a\Lambda}{\Lambda - \Delta_a\Lambda} = \frac{-(\Delta_0\Lambda - BC)}{\Lambda - \Delta_0\Lambda + BC}, \tag{7}$$

where γ is the degree of dissociation of the electrolyte. If the equilibrium between ions and ion pairs conforms to the law of mass action, we have

$$\frac{1}{K_a} = \frac{(\Lambda - \Delta_0\Lambda + BC)C\gamma^2f^2}{BC - \Delta_0\Lambda}, \tag{8}$$

where K_a is the association constant of the electrolyte. If K_a is known, the value of $\Delta_a\Lambda$ is given by the equation

$$-\Delta_a\Lambda = -(\Delta_0\Lambda - BC) = \frac{\Lambda K_aC\gamma^2f^2}{1 - K_aC\gamma^2f^2}. \tag{9}$$

In the limit, the value of the activity coefficient, f, is given by the Debye-Hückel equation

$$-\log f = \beta\sqrt{C} \tag{10}$$

At higher concentrations, it is more precisely given by the equation

$$-\log f = \frac{\beta\sqrt{C}}{1 + \kappa a}, \tag{11}$$

where $\kappa = 0.3288C \times 10^4$ and a is the sum of the radii of the ions.[3] For an aqueous solution at 0.01 N the value of f^2, calculated for $a = 5$ A, is approximately 3 per cent smaller than that calculated on the basis of equation (10).

The value of f may be approximated fairly closely on the assumption that γ equals unity. The value of B, on the other hand, is best approximated by the trial-and-error method. It may, of course, be calculated by equating the expression for $1/K_a$ as given by equation (8) for two different concentrations and solving for B, but, if the data for either concentration are in error, the values of B will also be in error. The latter method may at times be employed in a rough approximation.

For many salts the value of B is small and values of K_a may be approximated fairly closely on the assumption that $B = 0$. The values of K_a so found will vary only by a few per cent at higher concentrations. If B is large, the value of K_a so found will decrease greatly with increasing concentration.

In any case, values of K_a must be calculated for different values of B, and that value of B is the most probable which yields a K_a which reproduces experimental $\Delta_0\Lambda$ values most nearly.

Conductance data are available for ten electrolytes that exhibit negative deviations from the simple Onsager equation. The limiting slopes of the $\Delta_0\Lambda$ versus C plots range from -32 for Me_4NBr to -286 for Bu_4NI.[2] For two of these salts data are available up to 0.01 N; for the remainder the measurements were not carried above $2\text{–}3 \times 10^{-2}$ N, but the deviations of four salts are sufficiently large to permit a fairly satisfactory analysis. For the remaining four salts the deviations are too small to permit a satisfactory determination of association constants.

In Table 1, Sections A–F, are presented the results of our calculations. Values of Λ_0, B, and K_a employed in the calculations are given at the head of each subtable. Concentrations are given in the first column, experimental values of $\Delta_0\Lambda$, the deviations from Onsager's equation, in the second column, and the differences between calculated and experimental values of $\Delta_0\Lambda$ in the third column. Underneath the last column appear the average deviations of the calculated from the experimental values of $\Delta_0\Lambda$ for each salt.

Three parameters are involved in the calculations: Λ_0, the ion size a, and the coefficient B of the Shedlovsky equation. For five of the salts Λ_0 was adjusted slightly upward by, at most, 0.02 Λ-unit. For one salt, Pr_4NI, it was decreased 0.093 Λ-unit. In the original evaluation of Λ_0 for this salt, extrapolation was carried out from 7.5×10^{-4} N. In dilute solutions, K_a is highly sensitive to errors in Λ_0.

In the present calculations we have assumed $a = 5$ A for all salts with the exception of Me_4NI, where we assume $a = 3$ A. The calculated value of K_a is not very sensitive to the value assumed for a. For example, in the case of Bu_4NI, the value of f^2 calculated on the basis of $a = 5$ A differs from that calculated by means of equation (10) by only 3 per cent at $C = 0.01$ N. Thus the error introduced into the value calculated from K_a is only 3 per cent.

With the exception of Me_4NI, the value of B is small for all salts. For these salts one obtains a rather close approximation to the measured values of $\Delta_0\Lambda$ by

assuming $B = 0$. In the case of Me_4NI, however, values of $\Delta_0\Lambda$ calculated on this basis differ widely from the experimental ones.

It may be noted that the K_a values employed in these calculations are in all cases those obtained for the highest concentrations for which measurements have been reported.

TABLE 1

COMPARISON OF CALCULATED AND EXPERIMENTAL VALUES OF $\Delta_0\Lambda$ FOR DIFFERENT ELECTROLYTES

$C \times 10^4$	$\Delta_0\Lambda$exp	$\Delta_0\Lambda$exp − $\Delta_0\Lambda$calc
A. Bu_4NI*		
$\Lambda_0 = 96.291$; $B = 25$; $K_a = 3.869$		
2.13809	0.084	+0.014
4.25416	.135	.000
8.52156	.252	− .010
17.0437	.504	+ .002
34.0819	.940	.000
68.6093	1.713	− .002
101.735	2.358	0.000
	a.d. =	0.004
B. Me_4NI†		
$\Lambda_0 = 121.476$; $B = 235$; $K_a = 3.153$		
1.46496	0.019	−0.001
2.95012	.040	+ .002
5.39209	.067	+ .002
9.26201	.096	− .009
15.1065	.164	+ .007
23.8422	.229	+ .006
38.3611	.305	− .002
67.2754	.412	+ .006
96.1991	0.439	0.000
	a.d. =	0.004
C. Pr_4NI‡		
$\Lambda_0 = 100.287$; $B = 0$; $K_a = 2.387$		
7.51895	0.164	0.000
10.6598	.234	+ .005
17.6650	.369	+ .002
30.4199	0.603	0.000
	a.d. =	0.002
D. i-Am_4NNO_3*		
$\Lambda_0 = 89.282$; $B = 0$; $K_a = 1.995$		
2.83507	0.045	−0.003
5.68994	.095	+ .001
11.4082	.185	+ .003
22.8073	.347	.000
37.6061	0.548	0.000
	a.d. =	0.002
E. n-Bu_4NNO_3‡		
$\Lambda_0 = 90.900$; $B = 20$; $K_a = 1.723$		
5.24798	0.067	+0.001
7.76340	.095	.000
11.1045	.138	+ .004
16.1205	.189	− .001
23.1355	.265	+ .001
35.8263	0.391	0.000
	a.d. =	0.002
F. Bu_4NBr†		
$\Lambda_0 = 97.470$; $B = 0$; $K_a = 1.342$		
1.36593	0.012	−0.005
2.70109	.038	+ .004
5.39813	.065	− .001
10.7507	.124	− .003
21.4209	0.242	0.000
	a.d. =	0.003

* R. W. Martel and C. A. Kraus, these PROCEEDINGS, **41**, 9, 1955.
† E. L. Swarts and C. A. Kraus, these PROCEEDINGS, **40**, 382, 1954.
‡ E. G. Baker, thesis, Brown University, 1951.

The only test we have for the possible validity of the assumptions underlying these calculations in the closeness with which the calculated values of $\Delta_0\Lambda$ agree with those found experimentally. Since the value of f is only slightly affected by the value assumed for the parameter a, there are only two major assumptions. These are (1) that the ions of the electrolyte interact to form a neutral molecular species whose molecules are in equilibrium with the free ions in accord with the law of mass action and (2) that any deviations of the conductance of the free ions from Onsager's simple equation are positive and substantially linear up to the highest concentrations measured. As may be seen from Table 1, section A, the calculated and experimental values of $\Delta_0\Lambda$ for Bu_4NI are in rather good agreement. The average difference between the two is only ± 0.004 Λ-unit. For this average, two values have been included in the dilute range, where the differences are +0.014

and -0.010. Considering the numerous different sources of error inherent in the measurements of the conductance of electrolyte solutions, the agreement is remarkably good.

As may be seen from section B of Table 1, the agreement for Me_4NI is quite as good as or better than that for Bu_4NI. The concentration range is from 1.46×10^{-4} to 96.2×10^{-4}, a 60-fold concentration increase. The average difference between calculated and experimental values of $\Delta_0\Lambda$ is 0.004 Λ-unit, or slightly less than 0.004 per cent. The maximum difference is 0.009 Λ-unit. For the two salts for which we have data up to 0.01 N, the agreement is clearly within the limit of experimental error.

In Figure 1 are shown plots of $\Delta_0\Lambda$ and $\Delta_a\Lambda$ versus C for Me_4NI and Bu_4NI. Here the plots for $\Delta_0\Lambda$ represent the negative deviations from Onsager's equation as measured. The continuous line represents calculated values; the experimental values are represented by circles; the radii of the circles correspond to 0.05 Λ-unit.

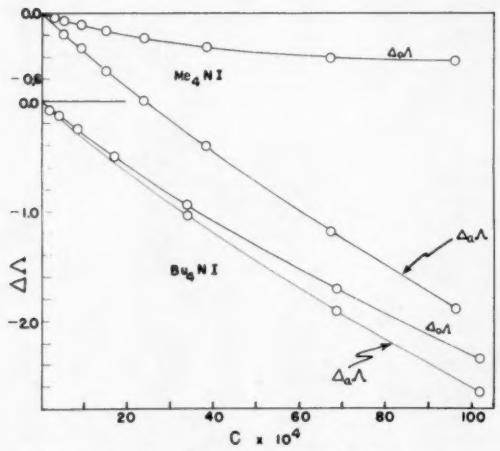


FIG. 1.—$\Delta_a\Lambda$ and $\Delta_0\Lambda$ as functions of concentration.

The plots of $\Delta_a\Lambda$ represent the conductance decrease due to ion association. The degree of association, $1 - \gamma$, is equal to $-\Delta_a\Lambda/(\Lambda - \Delta_a\Lambda)$. Obviously, a plot of $1 - \gamma$ versus C would parallel the plot of $\Delta_a\Lambda$ versus C fairly closely.

The values of K_a for the two salts, 3.153 (Me_4NI) and 3.869 (Bu_4NI), do not differ greatly; the two salts are associated to much the same degree, but the values of $\Delta_0\Lambda$ differ greatly. This is due to the fact that for the first salt $B = 235$ and for the second salt $B = 25$.

It was possible to determine B fairly closely for Bu_4NI, despite its small value. Assuming $B = 35$, the difference between calculated and experimental values of $\Delta_0\Lambda$ for four different concentrations, beginning with the highest, were respectively 0.000, -0.006, -0.002, and 0.000. As shown in Table 1, section A, assuming $B = 25$, the differences for the same concentrations are respectively 0.000, -0.002, 0.000, and $+0.002$. At high concentrations, a difference of 0.006 exceeds the experimental error. The values of $\Delta_0\Lambda$ as calculated for $B = 35$ are too large because the value assumed for B is too large.

At low concentrations the differences in the calculated values of $\Delta_0\Lambda$ on changing B from 25 to 35 are too small to be significant. It is only by means of data at concentrations above $C = 17.85 \times 10^{-4}$ that we are able to evaluate B with any degree of assurance. It is obvious, therefore, that, for the four salts for which data are available up to only about 3×10^{-3} N, it will not be possible to evaluate B successfully.

With Bu_4NNO_3 we get a good agreement using $B = 20$, but with Am_4NNO_3 we get an equally good agreement with $B = 0$. It would appear that, for the four salts here in question, B is small, probably less than 25. A precise evaluation of B is not possible on the basis of existing data.

The plots of $\Delta_a\Lambda$ for Me_4NI and Bu_4NI as shown in Figure 1, differ very little, for the K_a values are not greatly different and Λ_0 is markedly the greater for Me_4NI. The plots of $\Delta_0\Lambda$, however, differ greatly. For Bu_4NI the plot of $\Delta_0\Lambda$ differs little from that of $\Delta_a\Lambda$, since B has a value of only 25. For Me_4NI, on the other hand, B is equal to 235 and $\Delta_0\Lambda$ is very small. The differences between the ordinates of the $\Delta_a\Lambda$ and $\Delta_0\Lambda$ plots are equal to the values of BC.

For five of the salts the value of B is very small, not greater than 25. For Me_4NI the value of B is surprisingly large. Our knowledge of B values for different electrolytes is very limited. For KCl, KBr, and KI, Owen and Zeldes[4] report the following values at 25°: 145.0, 91.8, and 121.1, respectively. At 55°, for the same salts, they report 177.3, 184.0, and 267, respectively. Swarts and Kraus report B values as shown in Table 2. The $\Delta_0\Lambda$ versus C plots are all closely linear up to about 3×10^{-3} N, the highest concentration measured. It may be noted in this connection that the deviations reported by Owen and Zeldes are all closely linear, with the exception of KCl at 25°. It seems very likely that the deviations of the salts that we have examined in this paper are nearly linear, sufficiently so that any departure from linearity would not affect our results significantly.

TABLE 2

VALUES OF B FOR DIFFERENT SALTS IN WATER AT 25°

Salt	B	Salt	B
NaI	133	Me_4NBrO_3	67
NaPi	166	Bu_4NBrO_3	93
Me_4NPi	128	Me_4NCl	16
Et_4NPi	100	Me_4NNO_3	14

As Swarts and Kraus have shown, the negative deviations of quaternary ammonium salts decrease as we go from large to small positive ions; so, also, they decrease in going from iodide to chloride ions. It is of some interest to note that, in liquid HCN, where all deviations are positive, the positive deviations increase in the same order in which the negative deviations decrease in water.

The values of K_a which we have calculated are not in accord with Bjerrum's theory. According to this theory, association should increase as the size of the ions decreases. As our results show, the contrary is true; ion association decreases as the ions are smaller. This is clearly shown by the K_a values of Table 1. Most striking is the decrease in going from iodide to bromide ions; the association constant for Bu_4NBr is approximately one-third that of Bu_4NI. In this connection, it may be noted that for Me_4NCl, $B = +16$ and the $\Delta_0\Lambda$ versus C plot is strictly linear up to 33×10^{-4} N, the highest concentration measured. Thus it is highly probable that Me_4NCl is unassociated in water.

Bjerrum's equation, which relates the association constant of the ion-pair equilibrium with the dielectric constant and the temperature, contains a parameter a. This parameter may be interpreted as the distance between the centers of charge of the ions when in contact. For many electrolytes in solvents of lower dielectric constant for which K values have been determined on the basis of conductance

measurements, the parameter a has a value ranging from 4 to 8 A and increasing as the ions are larger. In other words, the K values are in reasonably good accord with Bjerrum's theory. In water, however, the a values derived from the K_a values of Table 1 are not in accord with this theory. For Bu_4NI the value found for a, calculated on the basis of $K_a = 3.87$, is only 1.02 A. Clearly, in this instance, the parameter a no longer has physical meaning. For Bu_4NBr, the value of a is found to be 1.93 A. In other words, the parameter a for the bromide is nearly twice that of the iodide.

We are presently carrying out an analysis of conductance data for solutions in dioxane-water mixtures. We shall report the results in a later communication.

* Fulbright scholar from the University of Rome. Postdoctoral fellow in chemistry at Brown University.

[1] T. Shedlovsky, *J. Am. Chem. Soc.*, **54**, 1905, 1932.

[2] E. L. Swarts and C. A. Kraus, these PROCEEDINGS, **40**, 382, 1954.

[3] H. S. Harned and B. B. Owen, *The Physical Chemistry of Electrolytic Solutions* (New York: Reinhold Publishing Corp., 1950), pp. 32–39, 120.

[4] B. B. Owen and H. Zeldes, *J. Chem. Phys.*, **18**, 1083, 1950.

THE OPTICAL ROTATORY DISPERSION OF SIMPLE POLYPEPTIDES. I

BY WILLIAM MOFFITT AND JEN TSI YANG

DEPARTMENT OF CHEMISTRY, HARVARD UNIVERSITY, CAMBRIDGE, MASSACHUSETTS

Communicated by G. B. Kistiakowsky, June 4, 1956

In a recent communication[1] the optical rotatory powers of helical macromolecules were investigated. It was shown that, under favorable circumstances, an abnormal frequency dependence should lead to marked departures from the simple Drude equation.

The purpose of the present note is twofold. In the first place, we shall propose an alternative phenomenological equation to describe the more complex dispersion that is sometimes observed for helices. The equation involves three constants, two of which should be more or less invariant properties of the helical skeleton itself; the latter should be independent of temperature, solvent, and the nature of the side chains for a given skeletal conformation. We shall then describe the observed dispersion for poly-γ-benzyl-L-glutamate and for poly-α-L-glutamic acid, a preliminary account of which has been published elsewhere.[2, 3] There is strong experimental evidence that, under the right conditions, both polymers exist in the α-helical configuration;[4, 5] it also seems clear that the screw senses of the helices are well defined, though whether these are right- or left-handed for given side chains remains to be established. The rotatory dispersion curves are anomalous in both cases, as had been predicted. Their more detailed analysis is also in satisfactory accord with the phenomenological equation proposed as a result of the theory.

In a forthcoming note,[6] the molecular theory is applied in greater detail to the α-helix. On the basis of dichroic studies of myristamide by Peterson and Simpson,[7]

it is suggested that the experimental evidence presented here is consistent with the view that the screw sense of both helices is right-handed.

ANALYSIS OF THE DISPERSION

It is profitable, at first, to consider why so many systems which are transparent through the near ultraviolet should show the simple dispersive behavior of a single-term Drude equation. In order to facilitate comparisons with the formulas of molecular optics, we shall begin by defining the most convenient quantities with which to work.

Instead of the specific rotation $[\alpha]$, it is preferable to use what may be called "effective molecular rotations" $[m']$ defined by the relations

$$[m'] = \left(\frac{3}{n_\nu^2 + 2}\right)[m] = \left(\frac{3}{n_\nu^2 + 2}\right)\frac{[\alpha]M}{100}. \tag{1}$$

Here n_ν is the refractive index of the medium, the dispersion of its polarizability for light of different frequencies ν being thereby eliminated. M is the molecular weight of the optically active solute species; for polymers it is the molecular weight per residue, when $[m']$ is called the "effective residue" or monomer rotation.

If the asymmetric molecules are in their ground states and interact neither with themselves nor with the solvent, then the frequency dependence of the effective molecular rotation is given by the quantal formula[8]

$$[m'] = \sum_i \frac{a_i \nu^2}{\nu_i^2 - \nu^2}, \qquad a_i = \frac{96\pi N_0}{hc} r_i, \tag{2}$$

where r_i is a molecular parameter (the rotational strength) associated with the ith absorption band, which occurs at frequency ν_i in the electronic spectrum. Under these conditions, the identity[8]

$$\sum_i a_i = 0 = \sum_i r_i \tag{3}$$

holds. However, since measurements are almost always conducted under less idealized circumstances, equation (2) may be used only if it is acknowledged that the rotational strengths r_i of the various absorption bands are functions not only of the solute molecules but also of the temperature and of the solvent. Moreover, since asymmetric molecules may be expected to have asymmetric solvation shells, equation (2) cannot describe the dispersion through regions of incident light frequency where the solvent itself absorbs.

Now suppose that we have to analyze the dispersion of simple molecular systems which only begin to absorb at wave lengths appreciably shorter than those for which the rotatory power has been measured. It will always be possible formally to expand the terms of equation (2) in inverse powers of $(\lambda^2 - \lambda_0^2)$, where the wave length λ_0 is as yet undetermined:

$$[m'] = \frac{\sum_i a_i \lambda_i^2}{(\lambda^2 - \lambda_0^2)} + \frac{\sum_i a_i \lambda_i^2 (\lambda_i^2 - \lambda_0^2)}{(\lambda^2 - \lambda_0^2)^2} + O[(\lambda^2 - \lambda_0^2)^{-3}]. \tag{4}$$

If $\lambda^2 \gg \lambda_0^2, \lambda_i^2$, this may be expected to converge quite rapidly. Now let λ_0^2 be determined by the relations

$$a_0\lambda_0^4 = \sum_i a_i\lambda_i^4, \qquad a_0\lambda_0^2 = \sum_i a_i\lambda_i^2. \tag{5}$$

The coefficient of $(\lambda^2 - \lambda_0^2)^{-2}$ on the right-hand side of equation (4) vanishes. Accordingly, if this λ_0^2 is positive and appreciably less than λ^2, the dispersion is well described by the simple formula

$$[m'] = \frac{a_0\lambda_0^2}{\lambda^2 - \lambda_0^2}. \tag{6}$$

This is, of course, just a slightly modified form of the single-term Drude equation. Since the dispersion of the factor $3/(n_\nu^2 + 2)$ is relatively small, it is effectively identical with this equation, whose empirical coefficients a_0, λ_0 receive their interpretation from equations (5). The optical rotatory dispersion of a large variety of molecules, including simple amino acids, di- and tri-peptides, obey equation (6) over a wide range of frequencies.[9]

It is difficult to lay down useful conditions for the validity of the simple Drude equation. We may note that, from equation (5), the contributions of the lower excited states are heavily weighted in determining λ_0^2; if these states are fairly close together, and the sum of their rotational strengths is larger than any one of these, then it is clear that λ_0^2 satisfies the required inequalities, particularly since equation (3) will be at least approximately valid. Conversely, if the sum of the rotational strengths of the lower excited states is small with respect to any one, then we may expect the dispersion to be complex, and equation (6) will not be satisfied. It is just this possibility—which must be expected whenever a principal source of the rotatory power is the asymmetric disposition of identical chromophores within a molecule—to which we attribute the anomalous dispersion of certain tartrates and, in particular, of simple polypeptides in a helical configuration.

In a helix of any great length, allowed transitions, which alone contribute to the optical rotatory power, are polarized either parallel or perpendicularly to the screw axis. (By "allowed" we do not imply that the bands need be strong.) For each parallel band there is an associated perpendicular band, and conversely. Moreover, pairs of such associated bands are rarely separated by more than 100 A or so; indeed, owing to their width, they will generally overlap. A single monomeric chromophore, like the peptide bond of a single residue, will appear in the spectrum of the α-helix as just such an overlapping pair of bands. More generally, this will be true for all monomeric chromophores which are prevented, on steric grounds or by the interposition of insulating groups, from conjugative interactions with one another. For these important, longer-wave-length systems the rotational strength of the parallel component may be exceedingly large, but it is almost exactly compensated by the opposing rotational strength of its perpendicularly polarized partner. The partial effective monomer rotation due to such a band pair may therefore be written[1]

$$\begin{aligned}[m_i'] &= \frac{a_{\parallel i}\lambda_{\parallel i}^2}{\lambda^2 - \lambda_{\parallel i}^2} + \frac{a_{\perp i}\lambda_{\perp i}^2}{\lambda^2 - \lambda_{\perp i}^2} \\ &\approx \frac{a_i\lambda_i^2}{\lambda^2 - \lambda_i^2} + \frac{b_i\lambda_i^4}{(\lambda^2 - \lambda_i^2)^2},\end{aligned} \tag{7}$$

since $|a_{\|i} + a_{\perp i}| \ll |a_{\|i}|$, $\lambda^2 - \lambda_i^2 \gg |\lambda_{\|i}^2 - \lambda_{\perp i}^2|$, and

$$a_i = (a_{\|i} + a_{\perp i}) + b_i, \qquad b_i = (a_{\|i} - a_{\perp i})\Delta_i,$$
$$\lambda_i = {}^1/_2(\lambda_{\|i} + \lambda_{\perp i}), \qquad \Delta_i = \frac{(\lambda_{\|i} - \lambda_{\perp i})}{\lambda_i}. \tag{8}$$

Formally, we therefore adopt the modified equation

$$[m'] = \sum_i \frac{a_i\lambda_i^2}{\lambda^2 - \lambda_i^2} + \sum_i \frac{b_i\lambda_i^4}{(\lambda^2 - \lambda_i^2)^2}, \tag{9}$$

where relation (7) has been extended over all pairs of band systems; in so doing, however, we recognize that b_i may occasionally be very small but that for certain terms a_i and b_i are quite comparable. It is evident that, in the latter eventuality, the dispersion cannot be represented by a simple expression like equation (6) over any sizable range of wave lengths, except in very special circumstances.

A phenomenological form for the anticipated dispersion is obtained in much the same way as equation (6) is derived from equations (2). It takes the form

$$[m'] = \frac{a_0\lambda_0^2}{(\lambda^2 - \lambda_0^2)} + \frac{b_0\lambda_0^4}{(\lambda^2 - \lambda_0^2)^2}, \tag{10}$$

where the constants a_0, b_0, λ_0 are chosen so that

$$a_0\lambda_0^2 = \sum_i a_i\lambda_i^2,$$
$$b_0\lambda_0^4 = \sum_i [a_i\lambda_i^2(\lambda_i^2 - \lambda_0^2) + b_i\lambda_i^4], \tag{11}$$
$$\sum_i [a_i\lambda_i^2(\lambda_i^2 - \lambda_0^2)^2 + 2b_i\lambda_i^4(\lambda_i^2 - \lambda_0^2)] = 0.$$

It will be valid whenever λ_0^2, so determined, is positive and considerably less than λ^2. We note that the last relation involves the sixth powers of the wave lengths. In such systems, even more than for the more simple systems discussed earlier, the dispersion is heavily dependent on the long-wave-length absorption bands and is only slightly influenced by the bands lying far out in the ultraviolet. In the event that λ_0 lies in the region of the longer-wave-length spectrum, it may happen that $|\lambda_i^2 - \lambda_0^2|$ is small with respect to λ_i^2, and equations (11) may be replaced by the simpler relations

$$a_0\lambda_0^2 = \sum_i a_i\lambda_i^2,$$
$$b_0\lambda_0^4 = \sum_i b_i\lambda_i^4, \tag{12}$$
$$b_0\lambda_0^6 = \sum_i b_i\lambda_i^6.$$

According to the molecular theory underlying the statements made in this paragraph, the parameters b_i, and the wave lengths λ_i for which these parameters are sizable, should be intrinsic properties of the helical skeleton; they should be insensitive—as should the constants b_0, λ_0 which they, in large measure, determine—to environmental factors such as arise from changes in solvent, temperature, and the nature of the side chains, always provided that this helical skeleton is not destroyed.

We do not, of course, claim that this analysis is any more than formal. Any results based on it are not unique. But it is no more possible at the present time to give an operational prescription for the validity of the Drude equation (6) than it is to say when our equation (10) may be expected to hold. Less so, for if it is found that b_0, λ_0 are indeed invariant properties for a given helical skeleton, under modifications affecting a_0 appreciably, our interpretation of these parameters in terms of equations (11) or equations (12) is greatly strengthened.

EXPERIMENTAL RESULTS

a) Poly-γ-benzyl-L-glutamate (MW, 130,000).—At 20° C. this polypeptide exists as a helix in 1,2-dichloroethane, N,N-dimethylformamide, 1,4-dioxane, and chloroform. Its optical rotatory power in these solvents is in sharp contrast to that which the polymer displays when dissolved in dichloroacetic acid at the same temperature. In the latter solvent the intramolecular hydrogen bonds are destroyed and the molecular configuration assumes a randomly coiled form.

The specific rotation of polybenzylglutamate in dichloroacetic acid decreases monotonically from −10° at 7500 A to −65° at 3400 A; the dispersion is simple and obeys the Drude equation, with λ_0 in the neighborhood of 1900 A, throughout this range of wave lengths. In the "helical" solvent, ethylene dichloride, however, the specific rotation is around $+8^1/_2°$ at 7000 A, rises to a maximum value of about $+10^1/_2°$ at 5700 A, and then decreases markedly, attaining the value of −181° at 3130 A. The behavior in chloroform, dimethylformamide, and dioxane is also anomalous, showing the same qualitative features as the ethylene dichloride solution. However, $[\alpha]$ is solvent-dependent: at 4360 A it takes the values +11°, −11°, −3°, and +1° for these respective solutions.

Our results for the "helical" solvents were fitted to equation (10) in the following manner. The specific rotations $[\alpha]$ were first converted to effective monomer rotations $[m']$, using Sellmeyer's formula to obtain suitable extrapolations for the dispersion of the refractive indices $n\nu$. $[m'](\lambda^2 - \lambda_0^2)$ was then plotted as a function of $(\lambda^2 - \lambda_0^2)^{-1}$ for trial values of λ_0 until a straight line was obtained. The procedure is illustrated for the ethylene dichloride solution in Figure 1. It is seen that the curvature depends fairly strongly on λ_0, whose value is therefore determined as 2120 ± 50 A. The slope of the line determines b_0, and its intercept on the ordinate axis determines a_0. For the other solutions λ_0 was found to be 2120 A also, within the limits of experimental errors. The results for these are collected in Figure 2 (*solid lines*). They may be expressed analytically in terms of equation (10), using λ_0 = 2120 A, together with the parameters assembled in Table 1. It will be remarked that, in addition to the approximate constancy of the critical wave length, the slopes of the lines are almost the same for all four solutions: both b_0, λ_0 appear to be substantially independent of the nature of the solvent, provided the helical configuration is maintained. The solvent-dependent term is clearly a_0.

Solutions of polybenzylglutamate in 1:3 ethylene dichloride:dichloroacetic acid show a most interesting behavior.[2] At 21° C. the rotatory dispersion is normal and much the same as in pure dichloroacetic acid at the same temperature: the polypeptide is randomly coiled. On raising the temperature to 41° C., however, it has been shown that the viscosity increases markedly (and reversibly) and that the optical rotatory power becomes anomalous: the helix is now the stable form. At the

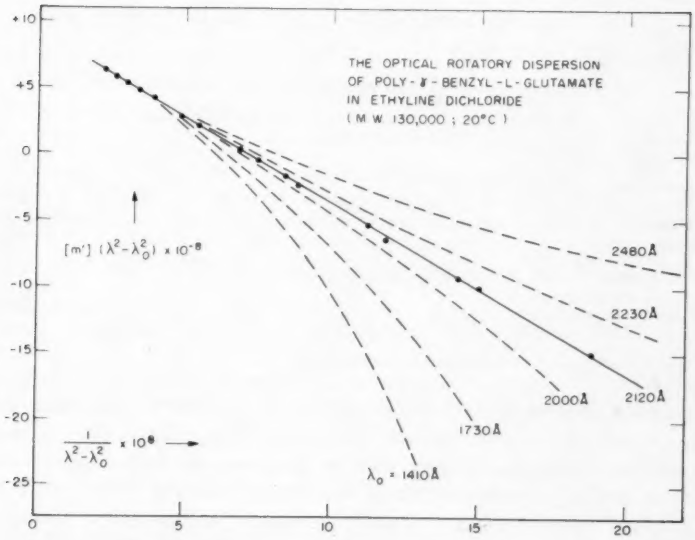

THE OPTICAL ROTATORY DISPERSION
OF POLY-γ-BENZYL-L-GLUTAMATE
IN ETHYLINE DICHLORIDE
(M.W. 130,000 ; 20°C)
[m'] (λ²-λ₀²) x 10⁻⁸
1/(λ²-λ₀²) x 10⁸
2480Å
2230Å
2120Å
2000Å
1730Å
λ₀ = 1410Å
+10
+5
0
-5
-10
-15
-20
-25
0
5
10
15
20

FIG. 1

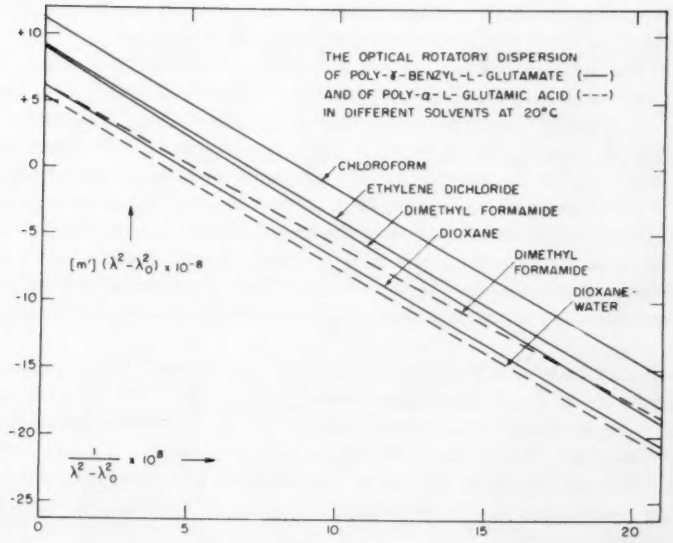

THE OPTICAL ROTATORY DISPERSION
OF POLY-γ-BENZYL-L-GLUTAMATE (——)
AND OF POLY-α-L-GLUTAMIC ACID (---)
IN DIFFERENT SOLVENTS AT 20°C
CHLOROFORM
ETHYLENE DICHLORIDE
DIMETHYL FORMAMIDE
DIOXANE
DIMETHYL FORMAMIDE
DIOXANE-WATER
[m'] (λ²-λ₀²) x 10⁻⁸
1/(λ²-λ₀²) x 10⁸
+10
+5
0
-5
-10
-15
-20
-25
0
5
10
15
20

FIG. 2

higher temperature the effective monomer rotation is very closely fitted by means of equation (10) with $\lambda_0 = 2120$ A, as before; b_0 is almost identical with that for solutions in pure ethylene dichloride at 20° C., although the intercept parameter a_0 is different. Their values are listed in Table 1.

TABLE 1*

PARAMETERS DESCRIBING THE OPTICAL ROTATORY DISPERSION (DEGREES)

	a_0	b_0
Poly-γ-benzyl-L-glutamate in		
Ethylene dichloride	205	−635
Dioxane	135	−630
Chloroform	250	−625
Dimethylformamide	200	−660
1:4 Ethylene dichloride:dichloracetic acid (44°C)	190	−630
Poly-α-L-glutamic acid in		
Dimethylformamide	135	−580
1:2 Dioxane:water (0.2 *M* NaCl, pH 4.72)	120	−625

* $\lambda_0 = 2120$ A in all cases; $T = 20°$ C. unless otherwise stated.

b) Poly-α-L-glutamic acid (M.W., 35,000).—Again at 20° C., polyglutamic acid exists as a helix in dimethylformamide. Its optical rotatory dispersion is anomalous in this solvent, changing sign near 4900 A to assume the large negative specific rotations at shorter wave lengths which characterize its benzyl ester. In 2:1 mixtures of 0.2 *M* aqueous sodium chloride:1,4 dioxane its behavior is strongly pH-dependent: at pH 6.56 the repulsions between the ionized carbonyl groups break down the helix and the configuration is that of an extended random coil; the rotatory dispersion is normal and closely follows the Drude equation with $\lambda_0 \approx 2100$ A. On acidifying to pH 4.72, however, the helical configuration is attained reversibly and the optical activity displays anomalous dispersion once more.

Both in the aqueous solution at pH 4.72 and in dimethylformamide, $[m']$ is found to obey equation (10) with $\lambda_0 = 2120$ A. On plotting $[m'](\lambda^2 - \lambda_0^2)$ against $(\lambda^2 - \lambda_0^2)^{-1}$, straight lines are again obtained for this critical wave length; these are shown as the broken lines in Figure 2. The parameters a_0, b_0 are listed in Table 1. The value of b_0 for the aqueous solvent falls in nicely with the similar values obtained for polybenzylglutamate, and that for the dimethylformamide solution is within 10 per cent of these: the slopes of the lines in Figure 2 are all much the same. It therefore appears that the constants b_0, λ_0 do not depend appreciably on the nature of the side chains attached to the α-helix. More data are, of course, required to establish this more firmly, but these must await the preparation of further polypeptides.

It has tacitly been assumed that the helical sense of both polypeptides is the same. Since the acid may be prepared from its benzyl ester in anhydrous benzene solution,[5] it seems unlikely that the α-helix should be destroyed and then rebuilt during the process. On forming the sodium salt, in aqueous solution, of the material so obtained, it attains the coiled conformation from which, however, it is possible to recover the acid in its helical conformation with no net change in its optical rotatory power.

The experimental details of the studies summarized in these paragraphs will be published elsewhere. The measurements ranged from 7500 A at the longer wave lengths down to 3340 A in the ultraviolet. It was found that, in each case, the

observed rotations over this range could be fitted by the parameters in Table 1 to within 4 per cent or to within the experimental error—whichever is the smaller.

Our samples were kindly provided by Dr. E. R. Blout. We are grateful to him and to Dr. P. Doty for stimulating discussions. The purchase of a Rudolph Model 200 S-80 spectropolarimeter was assisted by a generous grant from the William F. Milton Fund, Harvard University.

[1] W. Moffitt, *J. Chem. Phys.*, **24**, 1956.
[2] P. Doty and J. T. Yang, *J. Am. Chem. Soc.*, **78**, 498, 1956.
[3] P. Doty, A. Wada, J. T. Yang, and E. R. Blout, personal communication, not yet published.
[4] L. Pauling and R. B. Corey, these PROCEEDINGS, **37**, 241, 1951; P. Doty, A. M. Holtzer, J. H. Bradbury, and E. R. Blout, *J. Am. Chem. Soc.*, **76**, 4493, 1954; P. Doty, J. H. Bradbury, and A. M. Holtzer, *J. Am. Chem. Soc.*, **78**, 947, 1956; E. R. Blout and A. Asadourian, *J. Am. Chem. Soc.*, **78**, 955, 1956.
[5] E. R. Blout and M. Idelson, *J. Am. Chem. Soc.*, **78**, 497, 1956.
[6] W. Moffitt, these PROCEEDINGS, Vol. **42**, 1956 (in press).
[7] D. L. Peterson and W. T. Simpson, *J. Am. Chem. Soc.*, **77**, 3929, 1955.
[8] E.g., W. J. Kauzmann, J. E. Walter, and H. Eyring, *Chem. Revs.*, **26**, 339, 1940.
[9] E. Brand, E. Washburn, B. F. Erlanger, E. Ellenbogen, J. Daniel, F. Lippmann, and M. Scheu, *J. Am. Chem. Soc.*, **76**, 5037, 1954.

ISOTROPIC MATERIALS WHOSE DEFORMATION AND DISTORTION ENERGIES ARE EXPRESSIBLE BY SCALAR INVARIANTS

BY T. Y. THOMAS

DEPARTMENT OF MATHEMATICS, INDIANA UNIVERSITY

Communicated July 16, 1956

1. *Introduction.*—The deformation energy E per unit mass of an isotropic material with symmetric stress and rate-of-strain tensors given by σ and ϵ, respectively, is determined by the equation

$$\rho \frac{dE}{dt} = tr\ \mathfrak{d}\mathfrak{e}, \tag{1}$$

where ρ is the density and t denotes the time. If we define the deviation tensors σ^* and ϵ^* by writing

$$\sigma^*_{\alpha\beta} = \sigma_{\alpha\beta} - \frac{1}{3}\theta\delta_{\alpha\beta}; \qquad \epsilon^*_{\alpha\beta} = \epsilon_{\alpha\beta} - \frac{1}{3}\,tr\ \mathfrak{e}\ \delta_{\alpha\beta}, \tag{2}$$

with $\theta = tr\ \mathfrak{d}$, then

$$tr\ \mathfrak{d}\mathfrak{e} = tr\ \mathfrak{d}^*\mathfrak{e}^* + \frac{1}{3}\,\theta\ tr\ \mathfrak{e}; \qquad \xi = \xi^* + \frac{1}{3}\,\theta^2, \tag{3}$$

where we have put $\xi = \sigma_{\alpha\beta}\sigma_{\alpha\beta}$ and $\xi^* = \sigma^*_{\alpha\beta}\sigma^*_{\alpha\beta}$. Constructing the relations

$$\rho \frac{de^*}{dt} = tr\ \mathfrak{d}^*\mathfrak{e}^*; \qquad \rho \frac{de}{dt} = \frac{1}{3}\,\theta\ tr\ \mathfrak{e}, \tag{4}$$

it follows from the first equation (3) that E can be represented as the sum of the functions e^* and e. In accordance with the usual interpretation, we consider e^* to be the energy per unit mass due to *distortion*, i.e., change of shape as distinct from change of volume, while the function e is the energy per unit mass associated solely with change of volume. We shall call e^* the *distortion energy*. Since $e = E - e^*$, the simultaneous existence of the energies E and e^* implies the existence of the energy e.

In a recent paper entitled "Deformation Energy and the Stress-Strain Relations for Isotropic Materials"[1] we investigated the conditions which must hold in order that the deformation energy E be expressible as a scalar invariant of the tensors σ and ϵ when we assume a system of stress-strain relations of the form

$$\frac{D\sigma_{\alpha\beta}}{Dt} = f_{\alpha\beta}(\rho, \sigma, \epsilon), \tag{5}$$

where the f's are the components of a tensor invariant under proper orthogonal transformations relating moving rectangular co-ordinate systems, and the quantities $D\sigma_{\alpha\beta}/Dt$ are the components of the *absolute time derivative*[2] of the stress tensor. In particular, we considered a special class of relations (5) characterized by the requirement that the quantities $f_{\alpha\beta}$ were linear in the components $\epsilon_{\alpha\beta}$ and at most quadratic in the $\sigma_{\alpha\beta}$ with coefficients depending on the density ρ. Such systems were called *preferred* for convenience of reference. It was shown that there is a single class of preferred systems with energy E of the form $\psi(\xi + k\theta^2)/\rho$ and at most two classes of physically significant preferred systems for which E has the form $\phi(\xi + k\theta^2)$, where k is a constant. Denoting these classes by "Case A," "Case B," and "Case C," respectively, the associated energies E are given by

$$E = \frac{1}{4\mu\rho}\left(\xi - \frac{\lambda\theta^2}{3\lambda + 2\mu}\right) \qquad \text{(for Case A)}, \tag{6}$$

$$E = \frac{1}{4\mu}\left(\xi - \frac{\lambda\theta^2}{3\lambda + 2\mu}\right) \qquad \text{(for Case B)}, \tag{7}$$

$$E = \frac{-K}{4\mu}\log\left[1 - \frac{1}{K}\left(\xi - \frac{\lambda\theta^2}{3\lambda + 2\mu}\right)\right] \qquad \text{(for Case C)}, \tag{8}$$

where λ, μ, and K are constants. Expressions (6) and (7) have the formal character of the expression for the deformation energy in the ordinary elasticity theory. Expansion of the right-hand member of expression (8) shows that the same is also true to a first approximation for Case C systems.

It is the purpose of this communication[3] to deduce the further restriction which must be imposed on the preferred systems admitting the energies (6), (7), and (8) in order that these systems will also have distortion energies e^* which are scalar invariants of the stress tensor σ. The precise results are given in the following sections and will not be stated here. However, it will be emphasized that the conditions on the Case C system reduce this system to a simple generalization of the Prandtl-Reuss equations of the theory of plasticity. *As a consequence, the Prandtl-Reuss equations appear essentially as a condition for the energies E and e^* to be expressible by scalar invariants (see last equation in Sec. 4).*

2. *Case A Systems.*—For this case the matrix f of the components $f_{\alpha\beta}$ is given by

$$\begin{aligned} \mathbf{f} = \{ & \lambda\beta_0\theta + \lambda\beta_1\theta^2 + (3\lambda + 2\mu)\beta_1\xi + [2\mu\lambda + \lambda\alpha_1\theta + \lambda(\alpha_2 + \beta_2)\theta^2 \\ & + \lambda\alpha_3\xi - (3\lambda + 2\mu)\beta_2\xi]\, tr\, \boldsymbol{\varepsilon} - [(3\lambda + 2\mu)\alpha_1 + (3\lambda + 2\mu)\alpha_2\theta + \lambda\alpha_3\theta]\, tr\, \boldsymbol{\delta}\boldsymbol{\varepsilon}\} \boldsymbol{I}/2\mu \\ & + [\beta_0 + \beta_1\theta - \left(\frac{1}{2} - \beta_2\theta\right) tr\, \boldsymbol{\varepsilon} - \alpha_3\, tr\, \boldsymbol{\delta}\boldsymbol{\varepsilon}]\boldsymbol{\delta} + (2\mu + \alpha_1\theta + \alpha_2\theta^2 + \alpha_3\xi)\boldsymbol{\varepsilon}, \quad (9) \end{aligned}$$

where $\boldsymbol{I}$ is the unit matrix and $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$ are the matrices of the components $\sigma_{\alpha\beta}$ and $\epsilon_{\alpha\beta}$, respectively. The quantities λ and μ in expression (9) are constants, while the α's and β's either may be constants or may depend on the density ρ. Now from equations (2), (3), (5), and (9) we have

$$\begin{aligned} \frac{D\sigma^*_{\alpha\beta}}{Dt} = & \left[\beta_0 + \beta_1\theta - \frac{1}{2}\, tr\, \boldsymbol{\varepsilon} + \left(\beta_2 - \frac{1}{3}\,\alpha_3\right)\theta\, tr\, \boldsymbol{\varepsilon} - \alpha_3\, tr\, \boldsymbol{\delta}^*\boldsymbol{\varepsilon}^*\right] \sigma^*_{\alpha\beta} \\ & + \left[2\mu + \alpha_1\theta + \left(\alpha_2 + \frac{1}{3}\,\alpha_3\right)\theta^2 + \alpha_3\xi^*\right] \epsilon^*_{\alpha\beta}. \quad (10) \end{aligned}$$

Also, using the second relation (3), we observe that equation (6) can be written

$$E = \frac{1}{4\mu\rho}\left[\xi^* + \frac{2\mu\theta^2}{3(3\lambda + 2\mu)}\right]. \quad (11)$$

The separation of the quantities ξ^* and θ^2 on which the energy E depends, according to equation (11), suggests that we should have

$$e^* = \frac{\xi^*}{4\mu\rho}\,; \qquad e = \frac{\theta^2}{6(3\lambda + 2\mu)\rho}\,. \quad (12)$$

Substituting this expression for e^* into the first condition (4), making use of the equation of continuity

$$\frac{d\rho}{dt} + \rho\, \mathrm{tr}\, \boldsymbol{\varepsilon} = 0, \quad (13)$$

and equating to zero the coefficients of the different terms in the resulting equation, we find that

$$\beta_0 = \beta_1 = \alpha_1 = 0; \qquad \alpha_2 = -\frac{2\mu}{3K}; \qquad \beta_2 = \frac{2\mu}{3K}; \qquad \alpha_3 = \frac{2\mu}{K}, \quad (14)$$

where we have introduced the quantity K, which, in general, may be a function of ρ. Hence equation (10) becomes

$$\frac{D\sigma^*_{\alpha\beta}}{Dt} = 2\mu\left(1 + \frac{\xi^*}{K}\right)\varepsilon^*_{\alpha\beta} - \left(\frac{1}{2}\, tr\, \boldsymbol{\varepsilon} + \frac{2\mu}{K}\, tr\, \boldsymbol{\delta}^*\boldsymbol{\varepsilon}^*\right)\sigma^*_{\alpha\beta}. \quad (15)$$

From equation (5) and the trace of the matrix $\boldsymbol{f}$, we now find

$$\frac{d\theta}{dt} = (3\lambda + 2\mu)\, tr\, \boldsymbol{\varepsilon} - \frac{1}{2}\,\theta\, \mathrm{tr}\, \boldsymbol{\varepsilon}. \quad (16)$$

Relations (15) and (16) are equivalent to the stress-strain relations (5) with the f's given by equation (9) under the conditions (14). Moreover, since the second

relation (12) is an automatic consequence of relation (11) and the first relation (12), we have now proved the following result. *The stress-strain relations* (15) *and* (16) *admit the invariant energies E, e*, and e given by equations* (6) *and* (12). From the derivation of this result it is clear that equations (15) and (16) constitute the most general preferred system for which equations (6) and (12) give the energies E, e^*, and e.

3. *Case B Systems.*—The matrix f of the preferred system (5) admitting the deformation energy E given by equation (7) has the form

$$f = \{\lambda\beta_1\theta^2 - (3\lambda + 2\mu)\beta_1\xi + [2\mu\lambda + \lambda\alpha_1\theta + \lambda(\alpha_2 + \beta_2)\theta^2 + \lambda\alpha_3\xi - (3\lambda + 2\mu)\beta_2\xi]\, tr\, \boldsymbol{\varepsilon} - [(3\lambda + 2\mu)(\alpha_1 + \alpha_2\theta) + \lambda\alpha_3\theta]\, tr\, \boldsymbol{\delta\varepsilon}\} \frac{\boldsymbol{I}}{2\mu\rho} + (\beta_1\theta + \beta_2\theta\, tr\, \boldsymbol{\varepsilon} - \alpha_3\, tr\, \boldsymbol{\delta\varepsilon}) \frac{\boldsymbol{\delta}}{\rho} + (2\mu + \alpha_1\theta + \alpha_2\theta^2 + \alpha_3\xi) \frac{\boldsymbol{\varepsilon}}{\rho}, \tag{17}$$

where, as before, λ and μ are constants and the α's and β's are functions of the density ρ. Corresponding to the italicized result in Section 2, the following result can now be shown. *The system of stress-strain relations*

$$\frac{D\sigma^*_{\alpha\beta}}{Dt} = \frac{2\mu}{\rho}\left(1 + \frac{\xi^*}{K}\right)\varepsilon^*_{\alpha\beta} - \frac{2\mu}{\rho K}\, tr\, \boldsymbol{\delta}^*\boldsymbol{\varepsilon}^*\, \sigma^*_{\alpha\beta}, \tag{18}$$

$$\frac{d\theta}{dt} = \left(\frac{3\lambda + 2\mu}{\rho}\right) tr\, \boldsymbol{\varepsilon}, \tag{19}$$

where K may depend on the density ρ, *admits the deformation energy E given by equation* (7) *and also the energies*

$$e^* = \frac{\xi^*}{4\mu}; \qquad e = \frac{\theta^2}{6(3\lambda + 2\mu)}. \tag{20}$$

The system (18) and (19) is, of course, a special case of the system defined by equations (5) and (17) and is, moreover, the most general preferred system admitting the above energy invariants.

4. *Case C Systems.*—The most general preferred system admitting expression (8) for the deformation energy E is given by relation (5) and the matrix

$$f = \{\lambda\beta_1\theta^2 - (3\lambda + 2\mu)\beta_1\xi + [2\mu\lambda + \lambda\alpha_1\theta + \lambda(\alpha_2 + \beta_2)\theta^2 + \lambda\alpha_3\xi - (3\lambda + 2\mu)\beta_2\xi]\, tr\, \boldsymbol{\varepsilon} - [(3\lambda + 2\mu)(\alpha_1 + \alpha_2\theta) + \lambda\alpha_3\theta]\, tr\, \boldsymbol{\delta\varepsilon}\} \frac{\boldsymbol{I}}{2\mu\rho} + (\beta_1\theta + \beta_2\theta\, tr\, \boldsymbol{\varepsilon} + \beta_3\, tr\, \boldsymbol{\delta\varepsilon}) \frac{\boldsymbol{\delta}}{\rho} + (2\mu + \alpha_1\theta + \alpha_2\theta^2 + \alpha_3\xi) \frac{\boldsymbol{\varepsilon}}{\rho}, \tag{21}$$

in which

$$\alpha_3 + \beta_3 = -\frac{2\mu}{K}.$$

Here K is a constant, i.e., it is independent of ρ. The quantities λ and μ are again constants, while the α's and β's may be taken to be functions of ρ subject to the above relation in α_3 and β_3 by which the constant K is introduced. From equation

(21) we now have

$$\frac{D\sigma^*_{\alpha\beta}}{Dt} = \left[\beta_1\theta + \left(\beta_2 + \frac{1}{3}\beta_3\right)\theta\ tr\ \boldsymbol{\varepsilon} + \beta_3\ tr\ \mathfrak{d}^*\boldsymbol{\varepsilon}^*\right]\sigma^*_{\frac{\alpha\beta}{\rho}} + \left[2\ \mu + \alpha_1\theta + \left(\alpha_2 + \frac{1}{3}\alpha_3\right)\theta^2 + \alpha_3\xi^*\right]\frac{\varepsilon^*_{\alpha\beta}}{\rho}. \tag{22}$$

The form

$$E = \frac{-K}{4\mu}\log\left[1 - \frac{1}{K}\left(\xi^* + \frac{2\ \mu\theta^2}{3(3\ \lambda + 2\ \mu)}\right)\right] \tag{23}$$

of relation (8) suggests that

$$e^* = f(Z), \qquad Z = 1 - \frac{\xi^*}{K}. \tag{24}$$

Making this substitution in the first equation (4), we can deduce the following relation:

$$-\frac{f'}{K}\left\{2\left[\beta_1\theta + \left(\beta_2 + \frac{1}{3}\beta_3\right)\theta\ tr\ \boldsymbol{\varepsilon} + \beta_3\ tr\ \mathfrak{d}^*\boldsymbol{\varepsilon}^*\right]\xi^* + 2\left[2\ \mu + \alpha_1\theta + \left(\alpha_2 + \frac{1}{3}\alpha_3\right)\theta^2 + \alpha_3\xi^*\right]tr\ \mathfrak{d}^*\boldsymbol{\varepsilon}^*\right\} = tr\ \mathfrak{d}^*\boldsymbol{\varepsilon}^*.$$

Since this relation must be satisfied identically in the variables θ, ξ^*, $tr\ \boldsymbol{\varepsilon}$, and $tr\ \mathfrak{d}^*\boldsymbol{\varepsilon}^*$, it follows that

$$\alpha_1 = \beta_1 = 0; \qquad \alpha_2 = -\frac{1}{3}\alpha_3; \qquad \beta_2 = -\frac{1}{3}\beta_3, \tag{25}$$

and

$$\frac{4\mu Z}{K}\frac{df}{dz} + 1 = 0. \tag{26}$$

Integration of this latter relation (26) now gives

$$e^* = -\frac{K}{4\mu}\log\left(1 - \frac{\xi^*}{K}\right), \tag{27}$$

where the constant of integration is chosen so that $e^* = 0$ for $\xi^* = 0$, i.e., for vanishing stress.

Using equations (25), the stress-strain relations can now be expressed in the form

$$\frac{D\sigma^*_{\alpha\beta}}{Dt} = \frac{2\ \mu}{\rho}\left(1 + \frac{\alpha\xi^*}{K}\right)\epsilon^*_{\alpha\beta} - \frac{2\ \mu}{\rho K}(1 + \alpha)\ tr\ \mathfrak{d}^*\boldsymbol{\varepsilon}^*\sigma^*_{\alpha\beta}, \tag{28}$$

$$\frac{d\theta}{dt} = \left(\frac{3\lambda + 2\ \mu}{\rho}\right)\left(1 - \frac{\xi^*}{K}\right)tr\ \boldsymbol{\varepsilon} - \frac{2\ \mu}{\rho K}\theta\ tr\ \mathfrak{d}\boldsymbol{\varepsilon}, \tag{29}$$

where α can be taken to be a function of the density.[4] Hence we have arrived

at the following result. *The stress-strain relations* (28) *and* (29), *in which* λ, μ, K *are constants and* α *is a function of the density* ρ, *constitute the most general preferred system admitting the energies* E *and* e^* *given by equations* (8) *and* (27). The energy e is given by the difference $E - e^*$ of the deformation and distortion energies.

If $K > 0$, we see that the quantity in parentheses appearing in the right-hand member of equation (23) cannot attain the value K, since this would lead to an infinite value of the deformation energy E. In particular, this implies that the quantity ξ^* cannot reach the value K, as can, in fact, be seen directly from relation (27). To overcome this difficulty, we consider the energy expressions (8) or (23) and (27) for $K > 0$ to be valid in the physical sense only for sufficiently small values of the stress, i.e., we regard these formulas as approximate, with range of application limited to displacements which do not deviate too greatly from the natural or unstrained state of the material. No such restriction need be imposed for negative values of the constant K.

Taking $\alpha = 0$ in equation (28), we have

$$\frac{D\sigma^*_{\alpha\beta}}{Dt} = \frac{2\mu}{\rho}\left(\epsilon^*_{\alpha\beta} - \frac{tr\ \mathfrak{d}^*\mathfrak{e}^*}{K}\ \sigma^*_{\alpha\beta}\right).$$

These relations are identical with the Prandtl-Reuss equations of the theory of plasticity, apart from the occurrence of the quantity ρ in the right members. Moreover, equations (28) admit the yield condition $\xi^* = K$ in the sense that this relation is left invariant under a motion governed by equation (28). The stress-strain relations (28) and (29) can therefore be regarded as applicable both for small stresses and for large stresses capable of producing the plastic state as characterized by the usual quadratic yield condition.[5]

[1] *J. Math. and Phys.*, to appear.

[2] T. Y. Thomas, "Kinematically Preferred Co-ordinate Systems," these PROCEEDINGS, **41**, 762–770, 1955.

[3] This investigation was conducted at the request of the Naval Research Laboratory and has been supported by the Office of Naval Research under Contract Nonr-908(09), NR 041 037, with Indiana University.

[4] No attention has been paid to the question of the dimensions of the various terms which make up the preferred system. In this connection we observe that the α in eq. (28) should be a quantity having no physical dimension, a condition which can be satisfied most readily by choosing α to be an absolute constant independent of the density.

[5] The statements in this paragraph refer only to certain formal aspects of the system (28) and (29) and are not to be construed as meaning that these equations are necessarily suitable for the description of plastic behavior.

CYTOLOGICAL DEMONSTRATION OF THE HELICAL STRUCTURE OF GIANT CHROMOSOMES

By William Hovanitz

CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA

Communicated by F. W. Went, July 11, 1956

Alverdes[1] as long ago as 1912 interpreted the bands of the giant chromosomes found by Balbiani[2] in *Chironomus* as the result of spiralization. Later, Kaufmann,[3] applying plant-chromosome techniques to *Drosophila* giant chromosomes, concluded that the discoid or discontinuous appearance of these chromosomes was an expression of inadequate technique and that a spiral structure was the more accurate. Nevertheless, the discontinuous structure appeared to fit genetic data more easily.

Evidence from embryology[4, 5] indicates that the bands or "disks" of the giant chromosomes of the Diptera have their origin in the turns of a helical (spiral) structure. In view of the difficulty of discerning connections between the bands, however, this evidence lacked the necessary proof that the fully formed chromosomes were themselves spiral; in fact, it had been postulated that the coils actually broke to form rings. Evidence of a cytological nature has also been given[6] that the bands of the giant chromosomes are composed of the turns of many small helices which together appear to be a band. These data have been marginal in optical resolution, however. Other evidence[7] of a cytological nature depends largely upon a characteristic resolution of the microscope which is of uncertain meaning. Serial sections of *Drosophila* chromosomes[8] studied with the electron microscope appear to indicate that the structure is that of a hierarchy of helices, which appears to be similar to that found with isolated chromosomes.[9]

In the course of studying *Drosophila* giant chromosomes for the purpose of attempting to detect connections between the bands, a portion of one chromosome was found which showed a considerable area in which the bands actually were connected as in a helix. Further study has revealed that these connections are not uncommon, but it is rare to find one as extensive as this. To illustrate the helical nature, the strand was photographed from top to bottom at different focal levels by changing the fine focus adjustment of the microscope. The series of prints obtained does indeed show that the bands of the chromosome are in reality turns of helices and that the bands of the diploid or paired chromosomes are in fact composed of regularly arranged turns of two helices.

The series of prints is shown in Figure 1. Starting with one focal level at one end and proceeding to the other end, the turns of the helices can be followed all around the circumference of the coils between points *a* and *b* on the chromosome. In prints *A–F* the strands are slanted in the reverse direction from those in prints *J–K*. The central section of the helix, prints *G–I*, shows portions in both directions, as would be the case were one actually looking at a true helix.

The groove separating the paired homologues of the diploid chromosome is seen slightly below the middle. At *H* the spiral of one half of the chromosome is indicated to be independent of the spiral of the other half. In Figure 2 the same chromosome is shown flattened further by compression between the slide and coverslip. Being in a flatter field, the strands are seen over most of their distance

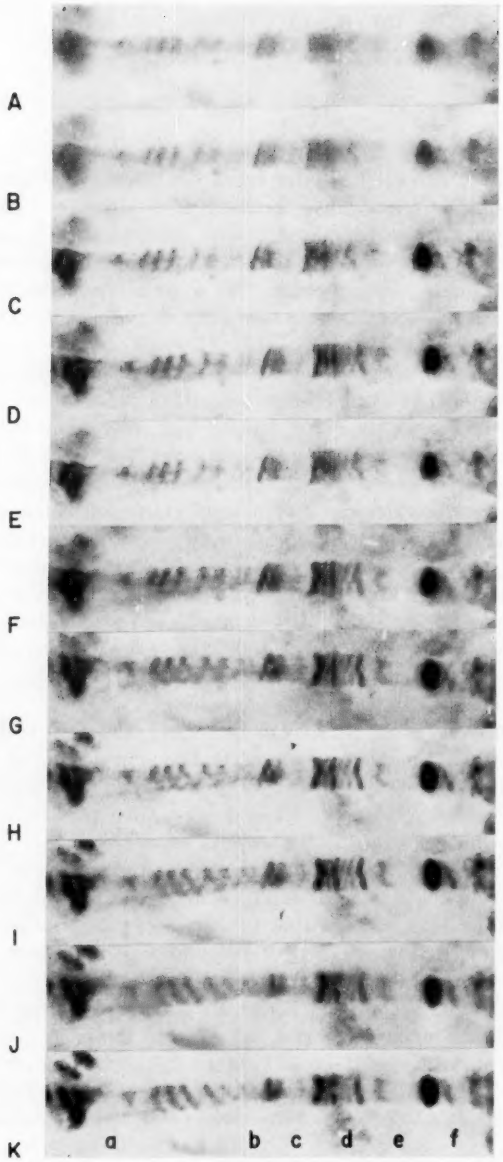


FIG. 1.—Part of a chromosome of *Drosophila melanogaster* (stained and fixed with aceto-orcein and smeared on a slide), photographed at successive focal-plane levels from top to bottom. Oil immersion obj. N.A. 1.32.

as they traverse a helical course and thus are seen to be proceeding in both directions at the same time. A model showing two helical wires representing the paired homologues is shown in Figure 3. Owing to the great depth of focus of the camera lens, all portions of the wire model are in focus at the same time, the lower parts having been shaded so as to indicate depth. However, in Figure 3, *B*, the upper portions of each coil are darkened; the appearance then is that of the chromosome pair in Figure 1, *A–F*. In Figure 3, *C*, the lower part of each coil is darkened and the appearance is that of the chromosome in Figure 1, *J–K*.

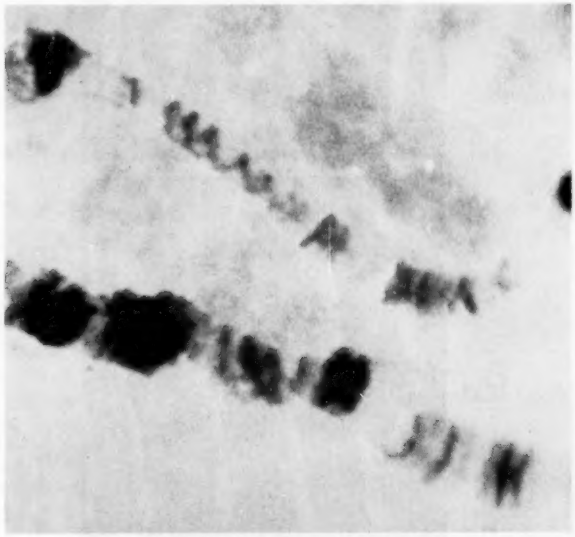


FIG. 2.—Same chromosome as in Figure 1 after a greater pressure on the coverslip. Note the further stretching between the bands.

From this, the larger bands of the diploid giant chromosomes are shown to be composed of the combined coils of the synapsed haploid strands. Furthermore, the bands are continuous from band to band, being connected together as a double helical structure, with the coils symmetrically arranged so that the "bands" appear to be paired. The close compression of the coils in the usual preparation makes it impossible to focus clearly on anything but the chromosome surface, which therefore looks like a series of disks or bands (Fig. 4, *a*). In the model of a single helical wire (Fig. 4), the distance between gyres is indicated to be either compressed (as at *a* or *e*) or in varying degrees of looseness (as at *b–d*). The loose areas show structure comparable with Figure 1 but fail to show the possible breakage of coils (as at *c*, *d*, or *e* in Fig. 1). A heavy band such as that at *e* in Figure 1 might arise as a tight concentration of coils as shown at *e* in Figure 4. However, there is another possible explanation to account for the apparent discontinuity between some bands, besides breakage, namely, the stretching of the strands so that they are below the

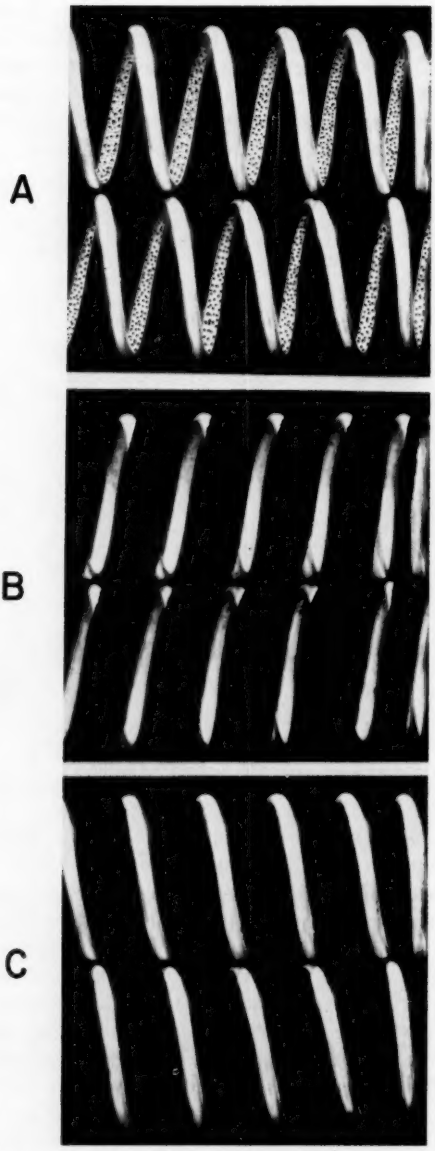


FIG. 3.—Metal models of helices arranged to show the appearance of a double helix (*A*), the same with the upper part of each coil blacked out (*B*), and the same with the lower part blacked out (*C*).

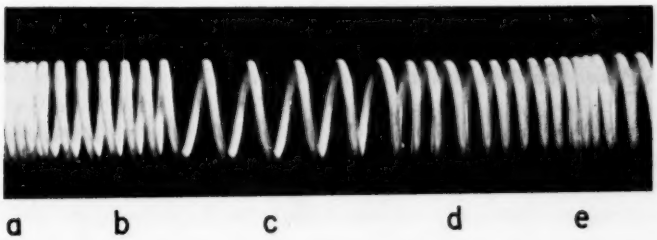


FIG. 4.—Model of a single helix with variable distances between the coils to illustrate the possible similarity to the giant chromosomes.

resolving power of the microscope and are thus invisible. This point will be brought out in a later paper that will consider the multiple-strand nature of the chromosome and of their hierarchy of helices.

The author wishes to express his gratitude to Dr. Taylor Hinton and Mr. W. Goodsmith of the Department of Zoölogy, University of California, Los Angeles, for their aid in supplying laboratory space, equipment, and the preparation shown.

[1] F. Alverdes, *Arch. Zellforsch.*, **9**, 168, 1912.
[2] E. G. Balbiani, *Zool. Anz.*, **4**, 637, 1881.
[3] B. P. Kaufmann, *Am. Naturalist*, **65**, 555, 1931.
[4] A. Sengün, *Rev. Fac. sci. Univ. Istanbul, B*, **19**, 169, 1954.
[5] R. Maria Albuquerque and J. A. Serra, *Mem. Est. Mus. Zool. Univ. Coimbra*, No. 209, pp. 1–9, 1951.
[6] H. Ris and H. Crouse, these PROCEEDINGS, **31**, 321, 1945.
[7] W. Hovanitz, *Wasmann J. Biol.*, **12**, 129, 1954.
[8] H. Gay, *Dissertation Abstr. Univ. Penna.*, **15**, 899, 1955.
[9] W. Hovanitz, A. R. T. Denues, and R. M. Sturrock, *Wasmann Collector*, **7**, 233, 1949.

*HETEROCARYOSIS AND PROTOPLASMIC INCOMPATIBILITY IN NEUROSPORA CRASSA**

BY LAURA GARNJOBST AND JAMES F. WILSON

DEPARTMENT OF BIOLOGICAL SCIENCES, STANFORD UNIVERSITY, STANFORD, CALIFORNIA

Communicated by E. L. Tatum, June 28, 1956

Introduction.—The physiological basis of incompatibility between organisms is a subject of considerable interest. Although incompatibility reactions are known in less complex forms, they have been studied most intensively in the higher organisms—for example, the researches concerned with outbreeding in flowering plants and blood types in animals. Among the micro-organisms, certain incompatibilities described in the past are related to sex, e.g., the mating-type factors of some Protozoa and fungi. In contrast, incompatibilities recently described in strains of *Neurospora crassa* appear to be related primarily to the vegetative stages. This

type of incompatibility is manifested by the failure of certain strains of like mating type to form successful heterocaryons.

Several biochemical mutant strains of *N. crassa* differing in ability to form heterocaryons have been investigated genetically.[1, 2, 3] Isolates from crosses made in this laboratory with strains of the *CD* and *cd* heterocaryon genotypes provided suitable material for further study of heterocaryosis. Since hyphal fusion ordinarily is a prerequisite for the experimental production of heterocaryons, an answer to the following question seemed important. Do the genes which inhibit heterocaryosis in certain unisexual combinations affect some steps in the fusion process, or do they affect some subsequent stage? Microscopic observation of growing hyphae showed that hyphal fusions do occur between all pairs of isolates that fail to form stable heterocaryons, as well as between all pairs that form stable heterocaryons. In the former pairs, however, fusion between hyphae is followed by a series of visible changes whereby the cells involved are destroyed. It is the purpose of the present paper to describe this "incompatibility reaction."

Materials and Methods.—Isolates of the *inositolless* (37401, *inos*) and *riboflavinless* (Y30539, *rib-2*) strains obtained in the genetic analysis previously described[2, 3] were used for most of the microscopic observations. The heterocaryon genotypes of the isolates of each biochemical mutant strain are *CD*, *cD*, *Cd*, or *cd*. Pairs of isolates (*inos* + *rib-2*) which give wild-type rate of growth on minimal medium (het^+ response) have the same heterocaryon genotype, i.e., *CD* + *CD*, *cD* + *cD*, *Cd* + *Cd*, and *cd* + *cd;* those which give no growth or delayed erratic types of growth on minimal medium (het^- response) have unlike heterocaryon genotypes, i.e., *CD* + *cD*, *CD* + *Cd*, *CD* + *cd*, *cD* + *Cd*, *cD* + *cd*, and *Cd* + *cd*. Other strains and isolates will be mentioned in connection with the particular tests concerned.

Material for examination was prepared under sterile conditions as follows. The cultures to be tested were grown at 30° C. in Petri plates containing a thin layer of glycerol complete medium[4] with 2.5 per cent agar. When the mycelium from a central inoculation of conidia had reached about halfway toward the periphery of the plate, small squares were cut along the advancing hyphal front for transfer to slides. Previous to the transfer, a small rectangle of glycerol complete agar (2.5 per cent) from an uninoculated plate had been put on each slide within a moistened Petri plate. The squares of agar with hyphae from each of the two cultures to be tested were then placed at opposite ends of the rectangle, so that the hyphal tips pointed toward the central "bridge." Such slide preparations, six for each pair, were incubated at 34° C. until the hyphae growing toward one another on the surface of the agar were about 0.3 mm. apart. A drop of minimal medium and a cover slip were then added for observation under the microscope at room temperatures (20°–28° C.).

The best results were obtained when the agar bridge and squares with hyphae were equal in thickness. A narrow bridge facilitated tracing the fused hyphae to their origins and allowed diffusion of oxygen. To support the cover slip and reduce evaporation, small squares or strips of agar were added to one or both sides of the bridge.

Both phase-contrast and ordinary compound microscopes were used. A magnification of 645× was found to be adequate for the observations presented.

Results.—As a basis for this study, the wild type SY7A was examined first.

Many types of intrastrain fusions were observed from the time of contact until protoplasmic flow took place through the new opening. The types of fusions (Figs. 1 and 2) are quite similar to those described in *Neurospora* and in other incompletely septate fungi.[5, 6]

The interstrain hyphal fusions of the *het*$^+$ pair *CD inos* and *CD rib-2* (in $a + a$ or $A + A$ mating types), as well as the intrastrain fusions of each of these isolates, were found to be indistinguishable from the intrastrain fusions observed in the wild type, as might be expected. The same was true of the intrastrain fusions of each member of the *het*$^-$ pair of isolates, *CD rib-2* and *cd inos* (in $a + a$ or $A + A$ mating types, and in the reciprocals, *CD inos* and *cd rib-2*), and also of the interstrain fusions up to and including formation of an "interseptum." Thereafter, a series of visible, irreversible changes was initiated.

To facilitate a description of these changes, one hypha is termed the "attacking hypha" and the other the "quiescent hypha." However, in most of the fusions observed the quiescent hypha responded by forming a protuberance (Figs. 1 and 9), and both hyphae often were equally active, i.e., visibly advancing toward one another on the agar surface. The subsequent series of changes occurred in the following order.

1. Shortly after contact (from 2 to 30 minutes or more at room temperatures of from 25° to 28° C.) protoplasmic movement took place through the new opening. The flow in some instances lasted about 45 or 60 seconds, occasionally somewhat longer. However, a mere start of movement or no perceptible movement at all were also common.

2. With or without movement, the protoplasm in the vicinity of the fusion, characteristically in both hyphae, sooner or later became gradually more granular, then finely vacuolated (Fig. 6). The vacuoles increased in size, often becoming quite large, as shown in Figures 7, 8, and 9. At the same time the intervacuolar portion became dense (Fig. 9). The conspicuous, affected region varied in extent from one to several cells in each hypha and usually was sharply delimited by septa, each containing a distinct opaque disk or plug (Figs. 5, 6, and 10) in place of an opening. At other times two or three consecutive septa in the outer cells were plugged. When this occurred, the outermost cells frequently showed a delayed effect or a graded effect, i.e., less granulation and vacuolation distally. The plugs were most prominent at the height of the vacuolation.

3. Finally, the vacuoles faded from view, leaving the contents of the affected cells pale and disorganized. Many hours later shrunken protoplasmic remnants could still be seen within the old hyphal walls.

During the late stage (stage 3) of the incompatibility reaction, the normal cells immediately behind the outermost plugged septa often developed new growing points (Figs. 10 and 11). These points formed new branches or instead grew within the old cell wall, pushing aside the plug, now partly resorbed, as well as the disintegrating protoplasm. When the latter course was taken, two growing tips within the wall of the old quiescent hypha were observed to approach one another and fuse, thus reconstituting this hypha (Fig. 12). New hyphae growing within attacking hyphae discontinued growth or broke out through the wall before reaching the point of union. Figures 11 and 12 illustrate the two types of regeneration observed.

The important result of the fusion and interaction of the two protoplasms *appears*

to be that the affected region in each hypha is effectively destroyed. However, the methods employed in this study do not rule out the possibility that some of the components of the invading protoplasm may escape into the adjoining unaffected region.

Representatives of all other *het*$^+$ and *het*$^-$ pairs of isolates (*inos* + *rib-2*) previously described were inspected microscopically. The results obtained with all *het*$^+$ pairs were essentially the same; that is, the intra- and interstrain fusions observed were indistinguishable and like the intrastrain fusions of the wild type SY7A. The difference between the intra- and interstrain fusions, as described in detail for the *het*$^-$ pair *CD rib-2* and *cd inos*, also holds for all other *het*$^-$ pairs tested; that is, all interstrain fusions induced the incompatibility reaction. No perceptible difference was observed in the reaction between pairs of strains differing by one or by two heterocaryon genes.

In a previous paper[3] it was shown that the effect of the genes concerned with heterocaryosis is independent of the mutant genes *inos* and *rib-2* and of all other biochemical mutant genes examined. Thus it seemed logical to assume that the incompatibility reaction would take place following fusion between two isolates of the *same* biochemical mutant strain differing in heterocaryon genotype. This was actually found to be the case. Representatives of all types of *het*$^-$ pairs of isolates previously described were tested.

The results of heterocaryon tests (unpublished) made in our laboratory with certain biochemical mutants suggested that the wild types from which they were derived differ in heterocaryon genotype. These wild types are SY7A or SY4a and ST A or ST a (St. Lawrence strains).[7] Since heterocaryon formation of nutritionally wild-type strains cannot be determined by the usual methods, it was thought that a difference might be detected microscopically. To test this possibility, the following pairs of wild types were observed: (1) SY4a f_9 + Lindegren (L) 25a; (2) ST a + SY4a f_9; (3) ST a + L25a; (4) *pe*^{m} *fl a* (Y8743-21 (13-7), *microconidial fluffy*) + SY4a f_9; (5) *pe*^{m} *fl a* + ST a; (6) *pe*^{m} *fl a* + L25a; and (7) SY4a f_9 + ST a. Consistent results were obtained as follows: (*a*) All strains formed functional intrastrain fusions in all preparations. (*b*) In preparations including a St. Lawrence strain, the interstrain fusions were of two types—completed fusions initiating an incompatibility reaction and uncompleted fusions apparently resulting from failure to produce an opening at the point of contact. (*c*) All other pairs were found to be compatible according to the new test.

The mating-type factors *A* and *a* in *Neurospora* have also been called incompatibility factors.[8] Since the function of these allelic genes is not well understood,

---

FIGS. 1–12.—Photomicrographs of living, unstained hyphae of *Neurospora crassa* growing on nutrient agar. ×600. Arrows indicate fusions not obvious in photomicrographs.

FIGS. 1 and 2.—Intrastrain fusions.

FIGS. 3 and 4.—Interstrain fusions, *het*$^+$ pairs.

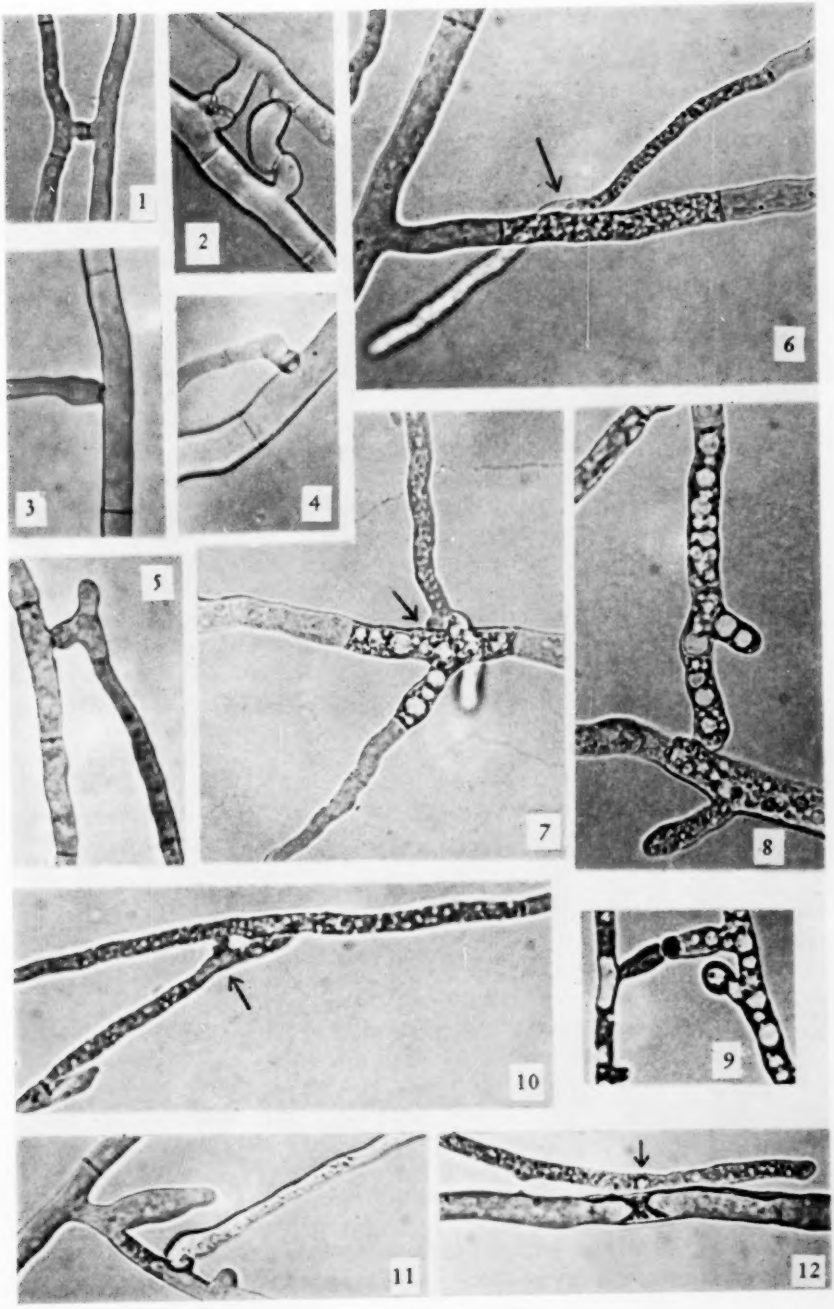
FIGS. 5 and 6.—Early stages of the incompatibility reaction, *het*$^-$ pairs. Plugs visible in septa delimiting affected regions.

FIGS. 7, 8 and 9.—Later stages of reaction showing prominent vacuolation: Fig. 7, in wild-type pair differing in mating type; Fig. 8, in pair differing in both mating type and heterocaryon genotype; Fig. 9, in pair differing in heterocaryon genotype.

FIG. 10.—Later stage of reaction in which vacuoles remained small. New growing point at lower left.

FIGS. 11 and 12.—Types of regeneration.

PLATE 1



it seemed of interest to extend the new test to strains differing in mating type. The pairs selected for these tests were *CD rib-2 a* + *CD inos A*, *CD inos a* + *CD inos A*, and *CD inos A* + *cd inos a*. In these tests interstrain fusions were infrequent in some slide preparations but common in other preparations of the same pair. Whenever the fusions did occur, an incompatibility reaction was initiated. The series of changes in most instances was similar in all respects to that described for *het*$^-$ pairs of the same mating type, but in the first two pairs listed an occasional difference was noted. This difference consisted of an unusually long period of protoplasmic flow through the interconnection, followed by a delay in the appearance of the reaction.

This unexpected evidence of incompatibility led to still further tests, this time with pairs of wild-type strains differing in mating type, as follows: (1) SY4a f_9 + SY7A; (2) SY7A + L25a; (3) SY7A + ST a; (4) ST A + ST a; and (5) ST A + L25a. As in the tests mentioned above, the interstrain fusions observed between these *A* and *a* strains also initiated an incompatibility reaction (Fig. 7). The variability described above was also found in pairs 1, 2, and 4. In addition, in pairs 3 and 5 a few fusions appeared to be uncompleted.

Although no constant difference in the incompatibility reaction was observed in *a* + *A* pairs as compared to *a* + *a* or *A* + *A* pairs, the results of the preliminary tests with strains differing in mating type suggest that the incompatibility reaction in these combinations is more readily affected by environmental conditions.

Discussion.—These findings raise certain questions regarding heterocaryosis, mating-type factors, and particularly regarding the nature of the incompatibility reaction. It has been concluded from a variety of experiments that functional anastomoses do not occur between mycelia of different species.[6, 9] The *Neurospora* species, *crassa*, *tetrasperma*, and *sitophila*, were among those tested.[6] The incompatibility reaction described in this paper, then, might be looked upon as similar to one of the many possible genetically controlled barriers leading to the evolution of sexual differentiation or speciation. This irreversible reaction appears to be an effective mechanism to prevent heterocaryosis in either unisexual (*het*$^-$) or bisexual[10, 11] combinations, but in the *A* + *a* combinations the stimulus inducing the reaction might be present only during the vegetative stages.

Numerous observations concerning hyphal fusion in fungi during vegetative growth and reproduction have been reported in the literature. Some of the findings described in this report are not new—for example, the open "pore"[12] in the center of each hyphal septum, the "pore-plug"[12] which appears as a result of injury to one or more cells in a hypha, and the growth of new hyphae within the old hyphal walls.[13] Most recently Gerdemann[14] described, in a phycomycetous fungus, a method of wound-healing which differs slightly from that observed in a number of species by Buller[12] and in *Neurospora* during the present study. These injury reactions resemble the incompatibility reaction but differ from it in one important respect—the absence of interhyphal fusion and subsequent intermingling of incompatible protoplasm.

In Protozoa there is at least one recorded observation of protoplasmic incompatibility not related to sex. Reynolds[15] investigated a "shattering reaction" in *Arcella polypora* and concluded that this reaction was a transitory change in phenotype induced by environmental differences. In contrast, the heterocaryon phenotypes of the *Neurospora* strains used in this study have remained consistent with their genotypes.

In comparing the *C* and *D* (or *c* and *d*) genes to the incompatibility factors of some higher fungi, as seems inevitable, it should be borne in mind that the strains of *Neurospora* obtained in the genetic analysis[2, 3] are either of the *a* or *A* mating type as tested with SY7A and SY4a (or ST A and ST a) and form perithecia when and only when the *a* and *A* strains are crossed, regardless of the heterocaryon genotype.

At this stage of the investigation in *Neurospora*, little is known of the relationship between the incompatibility reactions occurring (1) in pairs composed of strains of the same mating type and (2) in pairs composed of strains of different mating types. However, the results obtained with the various *het*$^+$ and *het*$^-$ pairs appear clear cut. The agreement between the heterocaryon tests previously made and the occurrence of the incompatibility reaction in all *het*$^-$ pairs is complete. Furthermore, the observations indicate that the cause of the incompatibility reaction lies in the protoplasmic portion of the mycelium.

Summary.—The genes *C* and *D* and the alternate alleles *c* and *d* in *N. crassa* are concerned with heterocaryon formation, as previously reported. Heterocaryon tests with pairs of isolates of the same mating type proved that all pairs of the same heterocaryon genotype form compatible heterocaryons, whereas all other pairs are incompatible, i.e., fail to grow or grow erratically on minimal medium. These two classes of pairs were termed *het*$^+$ and *het*$^-$, respectively.

The present experiments indicate that the reaction between members of all *het*$^-$ pairs involves protoplasmic incompatibility. Representative isolates of both *het*$^+$ and *het*$^-$ pairs of the *riboflavinless* (Y30539, *rib-2*) and the *inositolless* (37401, *inos*) strains were observed microscopically during vegetative growth, with results as follows: In the *het*$^-$ pairs all interstrain hyphal fusions initiated an irreversible incompatibility reaction. Upon exchange of protoplasm through the new opening between the hyphae, the cells in the vicinity of the fusion gradually disintegrated. Since no exception was found in any *het*$^-$ pair and all interstrain hyphal fusions in all *het*$^+$ pairs were normal, it is concluded that the incompatibility reaction is the mechanism whereby heterocaryotic growth in the *het*$^-$ pairs is inhibited or prevented.

* This work was supported in part by a grant from the Fleischmann Foundation.

[1] B. W. Holloway, *Genetics*, **40**, 117, 1955.
[2] L. Garnjobst, *Am. J. Botany*, **40**, 607, 1953.
[3] L. Garnjobst, *ibid.*, **42**, 444, 1955.
[4] E. L. Tatum, R. W. Barratt, N. Fries, and D. Bonner, *Am. J. Botany*, **37**, 38, 1950.
[5] E. Köhler, *Planta*, **8**, 140, 1929.
[6] E. Köhler, *ibid.*, **10**, 495, 1930.
[7] These strains were kindly supplied by Dr. Patricia St. Lawrence.
[8] E. R. Sansome, *Proc. Roy. Soc. Edinburgh, B*, **62**, 299, 1947.
[9] C. L. Taschdjian and E. Muskatblit, *Mycologia*, **47**, 339, 1955.
[10] G. W. Beadle and V. L. Coonradt, *Genetics*, **29**, 291, 1944.
[11] S. Gross, *Am. J. Botany*, **39**, 574, 1952.
[12] A. H. R. Buller, *Researches on Fungi*, Vol. **5**, Part I (Toronto: Longmans, Green & Co., 1931)
[13] B. O. Dodge, *Mycologia*, **12**, 115, 1920.
[14] J. W. Gerdemann, *Mycologia*, **47**, 916, 1955.
[15] B. D. Reynolds, *Biol. Bull.*, **46**, 103, 1924.

THE ORIENTATION AND EXTENT OF GENE TRANSFER IN ESCHERICHIA COLI

By P. D. Skaar* and Alan Garen†

BIOLOGICAL LABORATORY AND CARNEGIE INSTITUTION OF WASHINGTON, DEPARTMENT OF GENETICS, COLD SPRING HARBOR, NEW YORK

Communicated by M. Demerec, June 1, 1956

Introduction.—Genetic recombination of the type first described by Lederberg and Tatum[1] in *Escherichia coli*, strain K12, has been shown to involve conjugation[2] between sexually differentiated cells ($F+$ or *Hfr* with $F-$).[3, 4] Unilateral gene transfer from *Hfr* (high frequency of recombination) to $F-$ cells is suggested by the observation that recombinants occur only among the progeny of the $F-$ exconjugants.[2, 4] In recent experiments by Wollman and Jacob,[5] *Hfr*:$F-$ conjugants were separated by rapid stirring in a Waring Blendor, and then recombinants arising from the transfer and incorporation of two linked *Hfr* genes ($T+$ and $L+$) were selected by plating on appropriate minimal medium. It was found that, among these selected recombinants, the frequencies of several other, unselected *Hfr* genes increased with the duration of conjugation before the parental pairs were separated. These results indicate that all genes are not transferred simultaneously and suggest that, under these conditions, partial genomes may be transferred from *Hfr* to $F-$ cells. Further interpretation, however, is limited by the techniques employed. First, owing to the exclusive use of selected recombinants, the existence of a single, or preferred, sequence of gene transfer is not clear. It could be argued, for example, that during conjugation the probability of multiple gene transfer increases, but that no specific gene is regularly transferred earlier than another. In the present investigation the analysis of unselected recombinants has revealed that gene transfer is indeed oriented. Second, the method of disrupting pairs is sufficiently drastic to raise the possibility that the inferred transfer of partial genomes may not characterize "normal" mating.[6] In the present study this possibility has been minimized by the observation of exconjugants which originate in the absence of artificial disruption.

Materials.—Two *Hfr* strains are employed, both ultimately derived from the $F+$ strain, 58-161. W2323 is a subculture of the *Hfr* strain isolated by Hayes;[7] CS101 is a $V_1^r\ V_2^r$ derivative of W1895, the *Hfr* strain originally isolated by Cavalli.[8] CS2 is an $F-$ derivative of 58-161. W2323 and CS2 share the following genotype: $M-\ T+\ L+\ Th+\ Lac_1+\ Mal_1+\ V_1^s\ V_2^s\ V_6^s$; CS101 differs only in being $V_1^r\ V_2^r$. CS100 is an $F-$ strain derived from Y53 and is of the following genotype: $M+\ T-\ L-\ Th-\ Lac_1-\ Mal_1-\ V_1^s\ V_2^s\ V_6^r$. The origin of 58-161 and Y53 is described elsewhere.[9] The designations $M-$, $T-$, $L-$, and $Th-$ refer to methionine, threonine, leucine, and thiamine growth requirements, respectively, *Lac*− and *Mal*− to inabilities to ferment lactose and maltose, and V_1^r, V_2^r, and V_6^r to resistance to coliphages T1 (and T5), T2, and T6. Wild-type T1, T2H, and T6 were employed; T2 and T6 were inactivated by a dose of ultraviolet light which produced about 50 lethal hits per phage.

Experimental Procedures and Results.—When preincubated mixtures of CS101 (*Hfr Lac*+ *Mal*+) and CS100 ($F-$ *Lac*− *Mal*−) are plated on complete media

which permit the differentiation of parental fermentative types (EMB-lactose or EMB-maltose[10]), a high percentage of colonies are characteristically sectored. These colonies contain both fermenters and nonfermenters and resemble closely in appearance the segregating heterozygotes pictured by Lederberg.[11] Since these variegated colonies (*Lacv* and *Malv*) do not appear when preincubated mixtures of CS2 (*F*− *Lac*+ *Mal*+) and CS100 are similarly plated, they presumably reflect the mating process.

In order to examine the kinetics of formation of those units which give rise to variegated colonies on plating, the following standard procedures were adopted, which also met the demands of concurrent radiochemical studies.[12] Parental cultures, grown for 12 hours in Difco Penassay broth at 37° C. without aeration, were washed and resuspended in peptone broth (1 per cent Difco Bacto-Peptone, 0.3 per cent NaCl, 0.1 per cent glucose, 10^{-3} *M* $MgSO_4$, 10^{-4} *M* $CaCl_2$) at one-twentieth of their original concentration. After aeration at 37° C. for $2^1/_4$ hours (to give about 2×10^8 cells/ml), the cultures were centrifuged, washed, and resuspended at a twofold dilution in Penassay broth. After further aeration at 37° C. for 30 minutes, the cultures were centrifuged and concentrated fivefold in fresh Penassay broth (to give about 10^9 cells/ml). At this time, the cultures were mixed, 1 part *Hfr* to 2 parts *F*−, and aerated at 37° C. Under these conditions, the total cell titer increased about twofold during 80 minutes of mixed growth. After varying intervals, samples were removed from the mixture and either were directly diluted in saline and plated to give isolated colonies on EMB-lactose or EMB-maltose or were exposed to UV-inactivated T2 or T6 prior to dilution and plating. Phage exposure consisted of 5 minutes' adsorption in buffer at a multiplicity of 20 phage/cell and at a cell concentration of 2×10^8 per milliliter. The numbers of "+" (fermenters), "−" (nonfermenters), and variegated colonies were determined after 18–24 hours' incubation at 37° C. Mutant "+" papillae appeared later on "−" colonies and were easily distinguished from the more sectorial variegation.

At no time did variegated colonies (*Lacv* or *Malv*) appear following exposure to T2 (which attacks the *F*− parent), as is to be expected if gene transfer is unilateral from *Hfr* to *F*−.

Figure 1 shows the results of experiments in which measurements were made simultaneously of the frequencies of *Lacv* and *Malv* colonies before and after exposure to T6 (which attacks the *Hfr* parent) at different times during the mating period. Each point is based upon a sample containing 10^3–10^4 *Hfr* cells as pooled from three separate experiments. The total number of *Hfr* cells at each time was taken as the number of *Mal*+ and *Malv* colonies which appeared when cells were not exposed to T6 before plating. The results are consistent with the following interpretation.

Conjugating pairs yield variegated colonies (*Lacv* and *Malv*) when plated without exposure to T6. Exposure to T6 kills the *Hfr* parent and thus destroys the capacity of a pair to produce a variegated colony unless transfer of the pertinent gene to the *F*− cell has already occurred and will be followed by incorporation. In the latter event, since the surviving *F*− cell is multinucleate, it is still capable of producing a variegated colony. Therefore, *Lac*+ is transferred and incorporated more frequently than *Mal*+, since *Lacv* colonies suffer much less reduction than *Malv* colonies following exposure to T6. In fact, the number of *Malv* colonies after T6

exposure is so small that the number of *Malv* colonies which appear when cells are not exposed to T6 may be taken as a measure of plated pairs. The excess of *Lacv* over *Malv* observed when cells are not exposed to T6 is ascribable to fertilized $F-$ cells which have separated from their *Hfr* partners prior to plating. These "spontaneous" exconjugants appear as early as 10 minutes after mixing the parental strains and increase throughout the period studied.

On the preceding interpretation, exposure to T6 removes *Hfr* parental cells and leaves unfertilized $F-$ cells together with an unselected array of fertilized cells The latter are sufficiently numerous that their derivative clones may be analyzed by replica-plating.[13] Accordingly, mixtures of CS101 and CS100, after exposure to T6 at various intervals during the mating period, were plated on EMB-lactose (nonselective medium) as before. Then, after scoring for *Lacv*, all colonies (*Lac*−

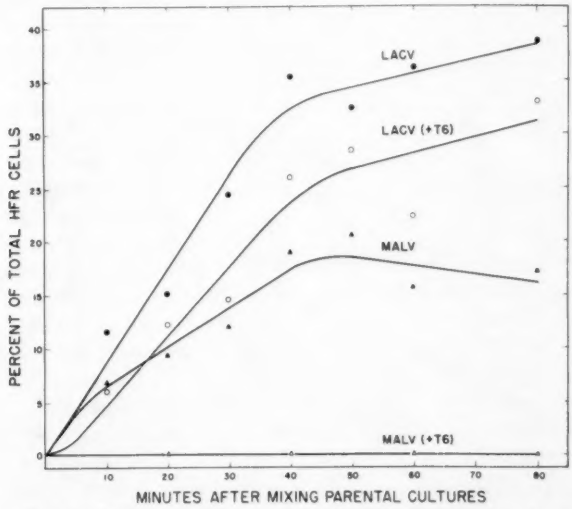


FIG. 1.—CS101 (*Hfr Lac*+ *Mal*+ V_6^s) × CS100 ($F-$ *Lac*− *Mal*− V_6^r). The frequencies of variegated colonies (*Lacv* and *Malv*) appearing on EMB-lactose or EMB-maltose when samples of the parental mixture were plated at different times. The curves designated as *Lacv* (+T6) and *Malv* (+T6) were obtained when mixtures were exposed to T6 before plating.

as well as *Lacv*) were examined for the presence of the *Hfr* genes, V_1^r and $L+$. V_1^r was detected on EMB-lactose spread with about 10^9 T1; $L+$ was detected on minimal medium (EMS-lactose[10]) supplemented with methionine, threonine, and thiamine (at concentrations of 10, 20, and 0.01 mg/l, respectively). The replica plates were scored after 18–24 hours' incubation at 37° C.

Colonies containing all seven possible combinations of the three *Hfr* genes scored (*Lac*+, V_1^r, and $L+$) were observed. Where two or three were present in a single clone, they were not invariably carried by the same subclone. Among those recombinant clones containing at least one of the three *Hfr* genes, three types predominated (94–99 per cent): Those which contained *Lac*+ but not V_1^r or $L+$, those which contained *Lac*+ and V_1^r but not $L+$, and those containing all three. As indicated in Figure 2 (*solid symbols*), the relative frequencies of these three

main types vary with the duration of mating prior to exposure to T6 and plating. This confirms the general conclusion of Wollman and Jacob that not all *Hfr* genes are transferred simultaneously. In addition, it provides evidence that transfer proceeds in a preferred sequence: *Lac* first, then V_1, and finally *L*.

The fact that this sequence of transfer is exactly opposite to the sequence suggested by the selected recombinants of Wollman and Jacob[5] requires explanation. The two sets of experiments differ in four important respects: the method of interrupting mating, the method of selecting recombinants, the *Hfr* strains used, and the lysogenicity of the strains used (Wollman and Jacob used nonlysogenic strains; all strains employed here are lysogenic for the temperate phage *lambda*). To determine the source of discrepancy in results, experiments were performed which used the T6-screening method described here and the lysogenic progenitor (W2323) of the *Hfr* strain employed by Wollman and Jacob. The results summarized in Table 1, although not extensive, appear to verify the temporal trend in gene frequencies reported by Wollman and Jacob. In particular, the fact that colonies containing *Lac*+ alone are always in a small minority, in contrast to the results obtained with CS101 (Fig. 2), indicates that differences between the *Hfr* strains (other than lysogenicity) are sufficient to explain the different sequences of transfer which have been observed. These differences may consist of structural inversions of the Lac_1—*L* region with respect to some point which is involved in the initiation of transfer. Jacob and Wollman[14] have discussed the possible significance of such structural differences in determining frequency of recombination.

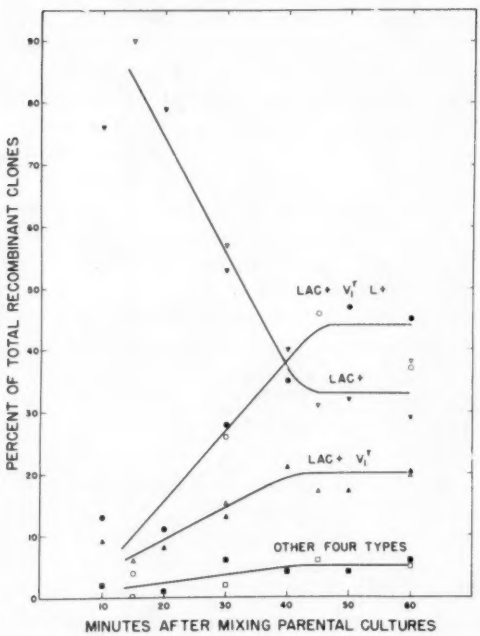


FIG. 2.—CS101 × CS100. The frequencies of different types of *F*− exconjugant clones obtained when samples of the parental mixture were plated at different times on nonselective medium. Each clone was scored for the presence of three *Hfr* genes: *Lac*+, V_1^r, and *L*+. "Total recombinant clones" is taken as the number which contained at least one of these genes. Solid symbols refer to clones obtained after T6 screening; open symbols refer to spontaneous exconjugant clones. At 10, 15, and 50 minutes, 50–100 clones were examined, at other times 100–300.

A reasonable model for the mating process has been afforded by observations on *Het* strain crosses (which yield persistent aneuploids).[11, 15] On this model, a complete *Hfr* (or *F*+) chromosome is transferred to a fertilized *F*− cell. Breakage then occurs at one or more *specific* points on this transferred chromosome. Reduc-

tion division follows rapidly, so that *Hfr* genes are lost unless crossovers have incorporated them into the complete *F*− chromosome. This explains the general predominance of *F*− markers in K12 crosses. The generality of this model has been questioned for some time,[16] but the experiments of Wollman and Jacob and ourselves provide the first strongly contrary evidence. This evidence consists of the fact that, to explain the temporal changes in gene frequencies, *varying* points of breakage of the *Hfr* chromosome must be assumed. If that part of the model based on *Het* crosses which calls for specific breakage points is not general, the generality of the part which calls for complete transfer of the *Hfr* genome is also open to question. The simpler alternative view that gene transfer may be normally partial becomes an economical working hypothesis.

TABLE 1

W2323 (*Hfr Lac* + *L* + V_6^s) × CS100 (*F* − *Lac*− *L*− V_6^r). Analysis of F-exconjugant clones containing either *Lac*+ or *L*+, obtained by exposure of the parental mixture to T6 after varying intervals of growth.

Clones Which Contained	Duration of Mixed Growth (Minutes) 15	30	45	60
Lac+ only	0	5	3	4
Lac+ and *L*+	5	31	43	58
L+ only	14	32	75	106
Total clones examined	19	68	121	168

However, in both the present experiments and those of Wollman and Jacob, exconjugants were obtained by artificial methods (exposure to phage or shearing in a blendor) which might induce abnormally partial transfer.[6] Therefore, an analysis was made of those exconjugants which arise spontaneously in a mating population of CS101 and CS100. This was done by plating samples at intervals directly onto EMB-maltose, where *F*− exconjugants and unmated *F*− cells form *Mal*− colonies. All colonies were then replica-plated to EMB-lactose, EMB-lactose plus T1, and EMS-lactose supplemented with methionine, threonine, and thiamine. Those *Mal*− colonies were then examined which contained one or more of the three *Hfr* genes, *Lac*+, V_1^r, and *L*+ (i.e., demonstrable exconjugants). The relative frequencies of combinations of *Hfr* genes among these exconjugant clones, plotted as open symbols in Figure 2, show that the same trend occurs whether or not mating is deliberately interrupted with phage. Therefore, discontinuities arise at varying sites in the *Hfr* genome under conditions corresponding closely to those prevailing in any bacterial cross, suggesting the normal transfer of partial genomes. In fact, the close agreement of trends among spontaneous and T6-induced exconjugants (Fig. 2) suggests that phage (and perhaps blending as well) may play no role in the *induction* of partial transfer but instead may screen for spontaneous exconjugants by rendering conjugating pairs incapable of producing recombinants.

Summary.—Nonsimultaneous gene transfer from *Hfr* to *F*− conjugating cells in *E. coli* K12 is confirmed. This transfer is oriented; certain genes regularly are transferred earlier than others. The sequence of transfer depends upon the *Hfr* strain employed, possibly due to structural differences. Sequential transfer is demonstrable among spontaneous exconjugants, strengthening the view that, normally, incomplete genomes are transferred.

The provision of bacterial strains by Dr. J. Lederberg is gratefully acknowledged, as are helpful discussions with Dr. A. D. Hershey.

* Part of this work was supported under contract by the United States Army Chemical Corps, Fort Detrick, Frederick, Maryland; part of the work was aided by a grant from the National Science Foundation.

† Fellow of the National Foundation for Infantile Paralysis. Aided by a grant (C2158) to Dr. A. D. Hershey from the National Cancer Institute of the National Institutes of Health, United States Public Health Service. Present address: Biophysical Laboratory, Purdue University, Lafayette, Indiana.

[1] J. Lederberg and E. L. Tatum, *Cold Spring Harbor Symposia Quant. Biol.*, **11**, 113–114, 1946.
[2] J. Lederberg, *J. Cell. and Comp. Physiol.*, **45** (Suppl. 2), 75–107, 1955.
[3] J. Lederberg, L. L. Cavalli, and E. M. Lederberg, *Genetics*, **37**, 720–730, 1952.
[4] W. Hayes, *J. Gen. Microbiol.*, **8**, 72–88, 1953.
[5] E. Wollman and F. Jacob, *Compt. rend. Acad. sci.* (Paris), **240**, 2449–2451, 1955.
[6] J. Lederberg, *Science*, **122**, 920, 1955.
[7] W. Hayes, *Cold Spring Harbor Symposia Quant. Biol.*, **18**, 75–93, 1953.
[8] L. L. Cavalli, *Boll. Ist. Sieroter. Milano*. **29**, 281–289, 1950.
[9] J. Lederberg, *Genetics*, **32**, 505–525, 1947.
[10] J. Lederberg, *Methods in Medical Research*, **3**, 5–22, 1950.
[11] J. Lederberg, these PROCEEDINGS, **35**, 178–184, 1949.
[12] P. D. Skaar and A. Garen, *Genetics*, **40**, 596, 1955.
[13] J. Lederberg and E. M. Lederberg, *J. Bacteriol.*, **63**, 399–406, 1952.
[14] F. Jacob and E. Wollman, *Compt. rend. Acad. sci.* (Paris), **242**, 303–306, 1956.
[15] T. C. Nelson and J. Lederberg, these PROCEEDINGS, **40**, 415–419, 1954.
[16] J. D. Watson and W. Hayes, these PROCEEDINGS, **39**, 416–426, 1953.

A PROPERTY OF SPECIAL JORDAN ALGEBRAS

BY A. A. ALBERT

UNIVERSITY OF CHICAGO*

Communicated June 29, 1956

Let $\mathfrak{M}$ be the set of all linear transformations on a finite-dimensional vector space over a field $\mathfrak{F}$ of characteristic not two. Then $\mathfrak{M}$ is an associative algebra with respect to the operation ab of the ordinary product of linear transformations a and b. But $\mathfrak{M}$ also forms a Jordan algebra $\mathfrak{M}^{(+)}$ with respect to the operation $a \cdot b = {}^1/_2(ab + ba)$.

At a recent Conference on Linear Algebras sponsored by the National Academy of Sciences–National Research Council, Nathan Jacobson conjectured that *every Jordan algebra $\mathfrak{B}$ is the homomorphic image of a special Jordan algebra $\mathfrak{A}$*, that is, the image of a subalgebra of a Jordan algebra $\mathfrak{M}^{(+)}$. In the present note we shall *disprove* the conjecture by showing that the known simple exceptional Jordan algebra is not such an image. Indeed, we shall derive the following more general result.

THEOREM. *Let $\mathfrak{B}$ be a Jordan algebra with radical $\mathfrak{H}$ such that $\mathfrak{B}$ is the homomorphic image of a special Jordan algebra $\mathfrak{A}$. Then $\mathfrak{B} - \mathfrak{H}$ is a special Jordan algebra.*

For $\mathfrak{A}$ is a Jordan subalgebra of an $\mathfrak{M}^{(+)}$ and has an enveloping associative subalgebra $\mathfrak{A}^*$ of $\mathfrak{M}$. Let $\mathfrak{N}$ be the radical of $\mathfrak{A}^*$, and form the semisimple associative image algebra $(\mathfrak{A}^*)' = \mathfrak{A}^* - \mathfrak{N}$. The elements of $(\mathfrak{A}^*)'$ are the cosets $x' = x + \mathfrak{N}$

defined for all elements x in $\mathfrak{A}^*$, and the mapping S, defined by

$$x \to x + \mathfrak{N},$$

is the natural homomorphism of $\mathfrak{A}^*$ onto $\mathfrak{A}^* - \mathfrak{N}$. Then S maps the special Jordan algebra $\mathfrak{A}$ onto a special Jordan subalgebra $\mathfrak{A}'$ of $[(\mathfrak{A}^*)']^{(+)}$ consisting of the cosets $a + \mathfrak{N}$ for a in $\mathfrak{A}$. It should be clear that

$$(\mathfrak{A}')^* = (\mathfrak{A}^*)'$$

is the enveloping associative algebra of $\mathfrak{A}'$. It is known[1] that the radical of $\mathfrak{A}$ is contained in the radical of $\mathfrak{A}^*$. Similarly, the radical of $\mathfrak{A}'$ is contained in the radical of $(\mathfrak{A}^*)'$, that is, $\mathfrak{A}'$ is a semisimple special Jordan algebra.

We are assuming that $\mathfrak{B}$ is the homomorphic image of $\mathfrak{A}$. Thus we may assume that $\mathfrak{B} = \mathfrak{A} - \mathfrak{C}$, where $\mathfrak{C}$ is an ideal of $\mathfrak{A}$. Let $\mathfrak{C}'$ be the subalgebra of $\mathfrak{A}'$ consisting of the cosets $c + \mathfrak{N}$ for c in $\mathfrak{C}$. Then $\mathfrak{C}'$ is an ideal of $\mathfrak{A}'$. Define

$$\mathfrak{B}' = \mathfrak{A}' - \mathfrak{C}'.$$

The elements of $\mathfrak{B}'$ are cosets $a' + \mathfrak{C}'$ for a' in $\mathfrak{A}'$. It is easy to verify that the mapping

$$a + \mathfrak{C} \to a' + \mathfrak{C}'.$$

is a homomorphism T of $\mathfrak{B}$ onto $\mathfrak{B}'$.

The kernel $\mathfrak{K}$ of T is the set of all elements $k + \mathfrak{C}$ of $\mathfrak{B}$ such that $k' + \mathfrak{C}' = \mathfrak{C}'$, that is, such that k' is in $\mathfrak{C}'$. The elements of $\mathfrak{C}'$ are the cosets $c + \mathfrak{N}$, where c is in $\mathfrak{C}$. Now $k' = k + \mathfrak{N}$ with k in $\mathfrak{A}$, and we have $k + \mathfrak{C} = h + \mathfrak{C}$, where h is in $\mathfrak{N}$. Thus k' is nilpotent, and $\mathfrak{K}$ is a nilideal. It follows that $\mathfrak{B}'$ is isomorphic to $\mathfrak{B} - \mathfrak{K}$, where $\mathfrak{K}$ is contained in the radical of $\mathfrak{B}$. However, $\mathfrak{B}'$ is the homomorphic image of a semisimple Jordan algebra $\mathfrak{A}'$ and so is semisimple. Hence $\mathfrak{K}$ must *be* the radical $\mathfrak{H}$ of $\mathfrak{B}$, and $\mathfrak{B} - \mathfrak{H}$ is isomorphic to $\mathfrak{B}'$. But $\mathfrak{A}'$ is semisimple and $\mathfrak{C}'$ is an ideal of $\mathfrak{A}'$, so that $\mathfrak{A}' = \mathfrak{C}' \oplus \mathfrak{D}'$, where, then, $\mathfrak{D}'$ and $\mathfrak{B}'$ are isomorphic. Hence $\mathfrak{B}'$ is isomorphic to the subalgebra $\mathfrak{D}'$ of a special Jordan algebra $\mathfrak{A}'$ and is special.

* This paper was sponsored in part by the Office of Ordnance Research, United States Army, under Contract No. DA-11-022-ORD-1571.

[1] Cf. Theorem 10 on p. 534 of the author's paper "On Jordan Algebras of Linear Transformations," *Trans. Am. Math. Soc.*, **59**, 524–555, 1946.

ON A CLASS OF FUNCTIONAL EQUATIONS OF MODULAR TYPE

By Richard Bellman

RAND CORPORATION, SANTA MONICA, CALIFORNIA

Communicated by H. S. Vandiver, July 3, 1956

1. *Introduction.*—The Voronoi functions, or generalized Bessel functions,

$$V_a(x, y) = \int_0^\infty e^{-\pi x^2 s - \pi y^2/s}\, ds/s^a, \tag{1.1}$$

play an important role in analytic number theory in connection with lattice-point problems and other investigations involving zeta functions. First introduced by Voronoi in his classic paper on the Dirichlet divisor problem, they have been treated by Hardy[1] and also by Steen,[2] who considered the more general functions defined by the relation

$$\int_0^\infty V(x; a_1, a_2, \ldots, a_k)x^{s-1}\, dx = \Gamma(s + a_1)\Gamma(s + a_2) \ldots \Gamma(s + a_k), \tag{1.2}$$

a class of functions whose importance was recognized by Voronoi. Extensions of these functions in the directions of algebraic number fields and matric fields have been sketched by Bellman,[3] and the matrix functions have been discussed in some detail by Bochner.[4] The function arises in the research of Hecke and in the continuation by Maass.[5]

In a previous paper[3] we sketched some extensions of the method used by Hardy to obtain a number of striking identities satisfied by series formed with these functions. In this paper our aim is to indicate how to form extensive classes of functions satisfying functional equations of modular type, using the function above and various extensions and generalizations within the rational field, algebraic number fields, matric fields, hypercomplex fields, and finite fields.

A more detailed account will be presented subsequently.

2. *The Rational Field.*—Consider the function defined by the series

$$f(x, y; u, v; t_1, t_2) = \sum_{m,\, n = -\infty}^{\infty} V_a((x + m)\sqrt{t_1}, (y + n)\sqrt{t_2})e^{2\pi i(mu + nv)}, \tag{2.1}$$

where $0 < x, y, u, v < 1$ and $t_1, t_2 > 0$. Using the representation in equation (1.1), this may be written as

$$f(x, y; u, v; t_1, t_2) = \int_0^\infty \left[\sum_{m=-\infty}^{\infty} e^{-\pi(x+m)^2 s t_1 + 2\pi i m u}\right] \left[\sum_{n=-\infty}^{\infty} e^{-\pi(y+n)^2 t_2/s + 2\pi i n v}\right] \frac{ds}{s^a}. \tag{2.2}$$

Applying the functional equation for the theta function in each of the brackets (in the previous paper[3] this was applied to only one of the brackets), we obtain the functional relation

$$f(x, y; u, v; t_1, t_2) = \frac{e^{-2\pi i(xu+yv)}}{\sqrt{t_1 t_2}} f(v, u; -y, -x; t_2^{-1}, t_1^{-1}).$$

3. *An Observation.*—Let us note that no properties of the weight ds/s^2 have been used. Consequently, the same functional equation holds for any weight function $dG(s)$ permitting the interchanges of summation and integration.

In addition, contour integrals of the form $\int_c e^{-\pi x^2 z - \pi y^2/z}\, dz/z^a$ may be used, with different contours yielding different functions, and slightly different types of identities.

4. *Specialization.*—If we allow x, y, u, and v to assume rational values, a number of interesting identities involving sums over divisors in specified residue classes are obtained.

This is a consequence of the relation

$$V_a(x, y) = x^{2a-2} \int_0^\infty e^{-\pi s - \pi(xy)^2/s} \frac{ds}{s^a} = x^{2a-2}\, V_a(|xy|), \qquad (x \neq 0). \quad (4.1)$$

Consider, in particular, the case where $a = 1$, and $x = y = u = v = 1/2$. Then equation (2.2) yields, with $t_1 t_2 = t^2$,

$$\sum_{m,\, n=-\infty}^{\infty} V_a\left(\frac{t|(1+2m)(1+2n)|}{4}\right)(-1)^{m+n} = -\frac{1}{t} \sum_{m,\, n=-\infty}^{\infty} V_a\left(\frac{|(1+2m)(1+2n)|}{4t}\right)(-1)^{m+n} \quad (4.2)$$

or

$$\sum_{R=1}^{\infty} V_a\left(\frac{Rt}{4}\right) a(R) = -\frac{1}{t} \sum_{R=1}^{\infty} V_a\left(\frac{R}{4t}\right) a(R), \quad (4.3)$$

where

$$a(R) = \sum_{|(1+2m)(1+2n)|=R} (-1)^{m+n}, \qquad -\infty < m, n < \infty.$$

Similar identities arise from other sets of rational values for x, y, u, and v.

5. *Generalizations in the Rational Field. I.*—Using the generalized Voronoi function in the form

$$V(x_1, x_2, \ldots, x_{k+1}; a_1, a_2, \ldots, a_k) = \int_0^\infty \cdots \int_0^\infty e^{-\pi s_1 x_1^2 - \cdots - \pi s_k x_k^2 - \pi x_{k+1}^2/s_1 s_2 \cdots s_k} \prod_{i=1}^{k} \frac{ds_i}{s_i^{a_i}} \quad (5.1)$$

cf. Steen[2] and Bellman,[3] a relation analogous to that of equation (2.2) can be obtained for the function defined by the series

$$\sum_{m_1=-\infty}^{\infty} V\left((x_1+m_1)\sqrt{t_1}, \ldots, (x_{k+1}+m_{k+1})\sqrt{t_{k+1}}; a_1, a_2, \ldots, a_k\right) e^{2\pi i \sum_{i=1}^{k+1} m_i u_i}, \quad (5.2)$$

where $0 < x_i, u_i < 1$, $t_i > 0$. These identities yield analogues of equation (4.3) for higher-order divisor functions.

6. *Generalizations in the Rational Field. II.*—Two further lines of generalization are immediate. In place of the two-dimensional version of equation (5.1), we may employ integrals of the type

$$\int_0^\infty \int_0^\infty e^{-\pi(s_1x_1^2+s_2x_2^2+s_1s_2x_3^2-x_4^2/s_1s_2)} \frac{ds_1 ds_2}{s_1^{a_1} s_2^{a_2}}, \tag{6.1}$$

utilizing all combinations of products, one at a time, two at a time, and so on. In place of forming sums such as that of expression (5.2), we can form sums of the type

$$\sum_{m_1=-\infty}^{\infty} V_a([Q((x_1+m_1),(x_2+m_2))]^{1/2}, (y+n)\sqrt{t_4}) e^{2\pi i(mu_1+m_2u_2+nu_3)}, \tag{6.2}$$

where $Q(u, v) = t_1u^2 + 2t_2uv + t_3v^2$ is a positive-definite quadratic form. Use of the multidimensional theta-function transformation will yield the desired functional equation.

7. *Algebraic Number Fields.*—Proceeding in the manner made classical by the work of Hecke, it is easy to form the corresponding Voronoi functions in totally real fields. Thus, for example, in $R(\sqrt{2})$, we have, as the analogue of the function in equation (1.1),

$$V_a(x, y) = \int_{s>>0} e^{-tr(\pi x^2 s + \pi y^2/s)} \frac{ds_1 ds_2}{(s_1^2 - 2s_2^2)^a}, \tag{7.1}$$

where the integration is over the region defined by $s_1 \geq \pm s_2\sqrt{2}$, and $x = x_1 + x_2\sqrt{2}$, $y = y_1 + y_2\sqrt{2}$.

8. *Matric Fields.*—Similarly, over matric fields, following the work of Siegel, we may define the function of two symmetric matrices X and Y,

$$V_a(X, Y) = \int_{S>0} e^{-tr(\pi X^2 S + \pi Y^2 S^{-1})} \frac{dS}{|S|^a}, \tag{8.1}$$

where the integration is over the region where the symmetric matrix S is positive definite, $dS = \prod_{1 \leq i \leq j \leq n} ds_{ij}$, and $|S|$ is the determinant of S (cf. Bellman[3] and Bochner[4]).

9. *Discussion.*—In a like fashion, generalized Bessel functions can be defined over a number of hypercomplex fields for which theta functions exist. This is an interesting direction of research which we shall develop in future papers.

10. *Finite Fields—Kloostermann Sums.*—An analogue of the Voronoi function in finite fields is the Kloostermann sum

$$K(x, q, p) = \sum_{n=1}^{p-1} e^{-2\pi i(x^2n+n^{-1})q/p}, \tag{10.1}$$

with a variety of other analogues obtainable from the continuous forms mentioned above. Results corresponding to those sketched earlier[3] and to those above can be obtained for these functions.

[1] G. H. Hardy, "Some Multiple Integrals," *Quart. J. Math.*, **39**, 357–375, 1908

[2] S. W. P. Steen, "Divisor Functions: Their Differential Equations and Recurrence Formulae," *Proc. London Math. Soc.*, **31**, 47–80, 1930.

[3] R. Bellman, "Generalized Eisenstein Series and Non-analytic Automorphic Functions," these PROCEEDINGS, **36**, 356–359, 1950. See also R. Bellman, "Wigert's Approximate Functional Equation and the Riemann Zeta-Function," *Duke Math J.*, **16**, 547–552, 1949.

[4] S. Bochner, *Harmonic Analysis and the Theory of Probability* (Berkeley: University of California Press, 1955).

[5] H. Maass, "Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirchletschen Reihen durch Funktionalgleichungen," *Math. Ann.*, **121**, 141–183, 1949.

ON THE PRINCIPLE OF INVARIANT IMBEDDING AND PROPAGATION THROUGH INHOMOGENEOUS MEDIA

BY RICHARD BELLMAN AND ROBERT KALABA

RAND CORPORATION, SANTA MONICA, CALIFORNIA

Communicated by Lyman Spitzer, Jr., June 15, 1956

1. *Introduction.*—In 1943, Ambarzumian[1] introduced a new approach to the study of atmospheric scattering problems. This technique was considerably extended by Chandrasekhar,[2] who utilized it to resolve a number of long-standing problems and gave it the name "principle of invariance."

In a series of papers, of which this is the first, we shall show that the full power of the method has not yet been realized. A number of problems of current interest, in scattering theory, in neutron diffusion, in the propagation of radio waves through stochastic media, as well as a variety of problems in other fields, may be treated by its use.

In addition to the varied physical applications, quite a number of mathematical questions arise concerning the relation between this approach and the classical techniques based upon partial differential and integral equations. These questions will be discussed subsequently.

In this note we are primarily interested in methodology. As an illustration of the general method, we shall discuss the scattering of light by an inhomogeneous plane medium of finite thickness. The corresponding problem for homogeneous media was treated by Ambarzumian and Chandrasekhar.

2. *Principle of Invariant Imbedding.*—Although space limitations do not permit us to discuss in any detail the close connections existing between the approach presented below and a number of fundamental techniques of classical analysis, we would like to mention some of these relations in a very brief fashion.

To begin with, the method is closely related to the "point of regeneration" method used in the study of stochastic processes, particularly in the study of "branching processes" arising in the study of bacterial growth, electron cascades, and cosmic-ray cascades (cf. Bellman and Harris;[3] S. Janossy[4]). Second, the invariance principle is intimately related to the principle of causality and hence to Huygens' principle (cf. Hadamard[5]). Consequently, there is a close connection with the theory of semigroups (cf. Hille[6]).

Finally, we should like to point out that the method is a decision-free version of the functional-equation technique of dynamic programming (cf. the "principle of optimality" [Bellman[7]]).

In his book[2] Chandrasekhar refers to an early use of the basic principle by Stokes.[8]

Let us now state the

PRINCIPLE OF INVARIANT IMBEDDING. *Given a physical system, S, whose state at any time t is specified by a state vector, x, we consider a process which consists of a family of transformations applied to this state vector.*

Suitably enlarging the dimension of the original vector by means of additional components, the state vectors are made elements of a space which is mapped into itself by the family of transformations. In this way we obtain an invariant process, by imbedding the original process within the new family of processes.

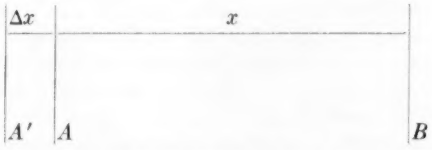
The functional equations governing the new process are the analytic expression of this invariance.

3. *Inhomogeneous Plane Media of Finite Thickness.*—To illustrate the method, let us consider one of the most immediate applications, the multiple scattering of light by an inhomogeneous plane medium of finite thickness. The homogeneous version of this problem was treated by Ambarzumian and Chandrasekhar.

In this case, the thickness of the slab will be the parameter describing the state of the process. It will be clear from what follows how the procedure may be modified to apply to the corresponding problems for spherical and cylindrical regions.

Consider the region bounded by two infinite parallel planes, possessing for the sake of initial exposition the requisite homogeneity properties as far as two space dimensions are concerned, so that it suffices to consider the plane region bounded by two parallel lines at a distance x apart.

In order to derive the equations governing the process, we shall, as usual, consider the area as consisting of a sequence of infinitesimally thin parallel zones, each possessing attenuation, scattering, and absorption parameters dependent upon the position of the zone:



The distance x is measured from the fixed plane B to the variable plane A.

Let parallel rays, the flux of radiation traversing unit length being πS, fall upon the boundary line A. Denote by u the cosine of the angle formed by the direction of the incident rays and the inward-drawn normal to A, and by v the cosine of the angle between the outward-drawn normal and a particular scattered ray. We wish to determine $Sr(v, u, x)$, the intensity of light reflected in the direction v. We assume that the ratio of scattering to absorption in the layer between x and $x + \Delta x$ is $\lambda(x)/(1 - \lambda(x))$.

The quantity $r(v, u, x + \Delta x)$ is the sum of five terms of order zero and one in Δx. All terms of smaller order are ignored. Following Ambarzumian,[1] we obtain

$$r(v, u, x + \Delta x) = r(v, u, x)\left(1 - \frac{\Delta x}{u} - \frac{\Delta x}{v}\right) + \frac{\lambda(x)\Delta x}{4v} + \frac{\lambda(x)}{2}\frac{\Delta x}{v}\int_0^1 r(w, u, x)\,dw + \frac{\lambda(x)}{2}\Delta x\int_0^1 r(v, z, x)\frac{dz}{z} + \lambda(x)\Delta x\int_0^1 r(w, u, x)\,dw\int_0^1 r(v, z, x)\frac{dz}{z}. \tag{3.1}$$

Letting $\Delta x \to 0$, this becomes

$$\frac{\partial r(v, u, x)}{\partial x} + \left(\frac{1}{u} + \frac{1}{v}\right) r(v, u, x) = \frac{\lambda(x)}{4v} + \frac{\lambda(x)}{2v}\int_0^1 r(w, u, x)\,dw + \frac{\lambda(x)}{2}\int_0^1 r(v, z, x)\frac{dz}{z} + \lambda(x)\int_0^1 r(w, u, x)\,dw\int_0^1 r(v, z, x)\frac{dz}{z}. \tag{3.2}$$

Introducing

$$r(v, u, x) = \frac{R(v, u, x)}{4v} \tag{3.3}$$

and performing a slight rearrangement, we obtain

$$\frac{1}{\lambda(x)}\left[\frac{\partial R(v, u, x)}{\partial x} + \left(\frac{1}{v} + \frac{1}{u}\right) R(v, u, x)\right] = 1 + \frac{1}{2}\int_0^1 R(w, u, x)\frac{dw}{w} + \frac{1}{2}\int_0^1 R(v, z, x)\frac{dz}{z} + \frac{1}{4}\int_0^1 R(w, u, x)\frac{dw}{w}\int_0^1 R(v, z, x)\frac{dz}{z}. \tag{3.4}$$

If $R(v, u, x)$ satisfies this equation, so also does $R(u, v, x)$. Taking for granted the uniqueness of solution, we write

$$\frac{1}{\lambda(x)}\left[\frac{\partial R(v, u, x)}{\partial x} + \left(\frac{1}{u} + \frac{1}{v}\right) R(v, u, x)\right] = \left[1 + \frac{1}{2}\int_0^1 R(v, z, x)\frac{dz}{z}\right] \times \left[1 + \frac{1}{2}\int_0^1 R(w, u, x)\frac{dw}{w}\right]. \tag{3.5}$$

The further substitution

$$\phi(v, x) = 1 + \frac{1}{2}\int_0^1 R(v, z, x)\frac{dz}{z} \tag{3.6}$$

yields

$$\frac{\partial R(v, u, x)}{\partial x} + \left(\frac{1}{u} + \frac{1}{v}\right) R(v, u, x) = \lambda(x)\,\phi(v, x)\,\phi(u, x). \tag{3.7}$$

Since $R(v,u,0) = 0$, we obtain

$$R(v, u, x) = \int_0^x e^{-(1/u + 1/v)(x-y)}\,\lambda(y)\,\phi(u, y)\,\phi(v, y)\,dy. \tag{3.8}$$

It follows that $\phi(u, x)$ satisfies the equation

$$\phi(u, x) = 1 + \frac{1}{2}\int_0^1 \left[\int_0^x e^{-(1/u+1/v)(x-y)}\, \lambda(y)\, \phi(u, y)\, \phi(v, y)\, dy\right]\frac{dv}{v}. \quad (3.9)$$

This leads to a feasible computational solution which we shall discuss in further detail subsequently.

[1] V. A. Ambarzumian, "On the Scattering of Light by a Diffuse Medium," *Compt. rend. Doklady Acad. sci. U.R.S.S.*, **38,** 257, 1943.

[2] S. Chandrasekhar, *Radiative Transfer* (London: Oxford University Press, 1950).

[3] R. Bellman and T. E. Harris, "On Age-dependent Binary Branching Processes," *Ann. Math.*, **55,** 280–295, 1952; "On Age-dependent Branching Processes," these PROCEEDINGS, **34,** 601–604, 1948.

[4] S. Janossy, *Cosmic Rays* (London: Oxford University Press, 1950).

[5] J. Hadamard, "Le Principe de Huygens," *Bull. Soc. Math. France*, **52,** 610–640, 1924.

[6] E. Hille, "Functional Analysis and Semi-groups," *Am. Math. Soc.*, 1948.

[7] R. Bellman, "The Theory of Dynamic Programming," *Bull. Am. Math. Soc.*, **60,** 503–516, 1954.

[8] G. G. Stokes, *Mathematical and Physical Papers*, Cambridge, 1904, Vol. IV: *On the Intensity of the Light Reflected from or Transmitted through a Pile of Plates*, pp. 145–156. Stokes' method was rediscovered independently by R. K. Luneberg (*The Propagation of Electromagnetic Plane Waves in Plane Parallel Layers* [Research Report No. 172-3, June, 1947, New York University, Washington Square College]), and R. M. Redheffer ("Novel Uses of Functional Equations," *J. Rat. Mech. Anal.*, **3,** No. 2, 271–279, 1954).

---

*ON MULTIPLICATIVE IDEMPOTENTS OF A POTENT SEMIRING**

BY SAMUEL BOURNE

UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA

Communicated by H. S. Vandiver, June 26, 1956

1. INTRODUCTION

A semiring is a system consisting of a set S and two binary operations in S called addition and multiplication such that (*a*) S together with addition is a semigroup; (*b*) S together with multiplication is a semigroup; (*c*) the left- and right-hand distributive laws $a(b + c) = ab + ac$ and $(b + c)a = ba + ca$ hold.

Semigroup is used in the sense of a closed associative system. R. Brauer mentioned to Vandiver[1] that Dedekind concerned himself with such an algebraic system without giving the semiring a formal definition. The latter was first introduced by Vandiver.[2] The simplest example of a semiring is the system of counting numbers relative to ordinary addition and multiplication. Vandiver also gave an example[1] of a semiring which cannot be imbedded in a ring. Another significant example is the semiring of finite and infinite cardinal numbers[3] which cannot be imbedded in a ring. Independently, this system came to us[4] naturally as the totality of endomorphisms $\mathfrak{E}$ of an arbitrary additive semigroup $\mathfrak{S}$.[5] A nonimbeddable semiring arises naturally when we take the totality of endomor-

phisms $\mathfrak{E}$ of a nonimbeddable additive semigroup. If both semigroups of a semiring are commutative, we say that the semiring is commutative. An important example of a commutative semiring is a distributive lattice.[6] In particular, the set of ideals of a commutative ring forms a commutative semiring relative to the operations sum and product.[7]

Our main purpose in this paper is to obtain an analogue for semirings of the fundamental lemma of structure theory in rings, that, if the left ideal $\mathfrak{l}$ of a ring $\mathfrak{R}$ with minimum condition is non-nilpotent, then $\mathfrak{l}$ contains a multiplicative idempotent.[8] Clifford[9] proved that in a semigroup $\mathfrak{S}$ without nilpotent ideals $\neq(0)$ and in which every two-sided ideal contains at least one left and at least one right minimal ideal of $\mathfrak{S}$, every left ideal $L \neq (0)$ of $\mathfrak{S}$ contains an idempotent element $\neq(0)$. We now show that this theorem remains true for semirings.

This paper has benefited materially from discussion and correspondence with A. H. Clifford, of Tulane University, and H. Zassenhaus, of McGill University.

2. ADDITIVE AND MULTIPLICATIVE IDEMPOTENTS

Definition 1: An additive identity, called *zero*, is an element 0, such that $0 + s = s + 0 = s$ for all s in S.

Definition 2: An additive idempotent of a semiring S is an element a such that $a + a = a$ and a multiplicative idempotent is an element m such that $m^2 = m$, $m \neq 0$.

A distributive lattice is an example[10] of a commutative semiring in which each element is both an additive and a multiplicative idempotent.

If S possesses a zero, then $a0 = a(0 + 0) = 2a0$ and $0a = 20a$ are additive idempotents which are not necessarily equal.

Example 1: Let $\mathfrak{S}_2$ be the noncommutative additive semigroup of order 2 with elements 0, 1 whose additive table is

+	0	1
0	0	1
1	0	1

Let $\mathfrak{E}$ be the totality of endomorphisms of $\mathfrak{S}_2$. Then $\mathfrak{E} = \{0, I, A, B\}$, where $(0, 1)0 = (0, 0)$, $(0, 1)I = (0, 1)$, $(0, 1)A = (1, 1)$, and $(0, 1)B = (1, 0)$. The addition and multiplication tables in the semiring $\mathfrak{E}$, are, respectively,

+	0	I	A	B
0	0	I	A	B
I	0	I	A	B
A	0	I	A	B
B	0	I	A	B

·	0	I	A	B
0	0	0	A	A
I	0	I	A	B
A	0	A	A	0
B	0	B	A	I

Since $0A = A \neq 0 = A0$, we have two distinct additive idempotents.

Example 2: Let S_2 be the semiring of order 2 with elements 0, 1 whose addition and multiplication tables are, respectively,

+	0	1
0	0	1
1	1	1

·	0	1
0	1	1
1	1	1

In this example S_2 possesses a zero element 0 in which $0^2 \neq 0$.

We shall assume that S possesses a zero element 0 and $0s = s0 = 0$, for all s in S.

Definition 3: A division semiring is a semiring in which the elements $\neq 0$ form a multiplicative group.

3. TWO-SIDED IDEALS CONTAINING MINIMAL RIGHT IDEALS

For the sake of completeness, we repeat the definition of *right ideal* given elsewhere.[4]

Definition 4: A right ideal S is a subset R of S containing zero such that if r_1 and r_2 are in R, then $r_1 + r_2$ is in R, and if r is in R and s is any element of S, then rs is in R.

Definition 5: The sum over a given set Δ of right ideals R_t is the smallest right ideal containing all the given right ideals. It consists of all finite sums $\sum r_t$, where each r_t belongs to one of the right ideals R_t. We denote this right ideal by $\sum R_t$. This addition is both commutative and associative in the widest sense. In the case addition, in the semiring S, is commutative, our definition of the sum of right ideals becomes identical with the standard definition in ring theory; hence $R_1 + R_2$ is the set of sums $r_1 + r_2$ with r_1 in R_1 and r_2 in R_2.

Definition 6: The product R_1R_2 of the right ideals R_1 and R_2 of a semiring S is the set of all finite sums $\sum r_{1i}r_{2i}$ with r_{1i} in R_1 and r_{2i} in R_2.

This multiplication of ideals is associative but not necessarily commutative. Also, the general distributive law $\sum R_i \sum Q_j = \sum_{i,j} R_iQ_j$ holds.

Let us mention the fact that the intersection of any number of ideals of the same kind is again an ideal of the same kind.

Definition 7: A right ideal N is said to be nilpotent if a power N^p is equal to the zero ideal (0).

Definition 8: A semiring S is said to be potent if it contains no nonzero nilpotent right ideals and left ideals.

LEMMA 1. *If R is a minimal right ideal of a semiring S, then tR, where t is in S, is either a minimal right ideal S or $tR = (0)$.*

Proof: tR is a right ideal. Let A be a right ideal contained in $tR \neq (0)$ and R_1 the set of all elements r_1 of R such that tr_1 is in A. Since $t(r_{1k} + r_{1e}) = tr_{1k} + tr_{1e}$ and $t(r_{1k})S = t(r_{1k})S$ are in A, we have that $r_{1k} + r_{1e}$ and $r_{1k}S$ are in R_1. This implies that $R_1 \neq (0)$ is a right ideal contained in R. Therefore, $R_1 = R$, $tR_1 = tR \subseteq A$, and $A = tR$. This proves that tR is minimal.

THEOREM 1. *If I is a two-sided ideal containing a minimal right ideal of S, the sum J of all minimal right ideals of S contained in I is a two-sided ideal of S. If I is minimal, then it is a sum of minimal right ideals of S.*

Proof: Since J is the sum of right ideals R_i, it is a right ideal. We show that it is also a left ideal. Let $j \in J$ and $s \in S$; then $sj = s\sum r_i = \sum sr_i$, where r_i belongs to some minimal right ideal R_i contained in I. Now $sr_i \in sR_i$, and sR_i is, by Lemma 1, a minimal right ideal contained in I, or $sR_i = (0)$. Hence, in either case, $sj \in J$ and J is a left ideal. Obviously, if I is minimal, $I = J$.

THEOREM 2. *If M is a minimal two-sided ideal of a potent semiring S, containing a minimal right ideal of S, then every minimal right ideal of S contained in M is a minimal right ideal of M.*

Proof: Let $R \neq (0)$ be a minimal right ideal of S contained in M, and A a right ideal $\neq (0)$ of M contained in R. Now AM is a right ideal of S, and $AM \subseteq RM \subseteq R$. Since R is minimal, this implies that $R = AM \subseteq A$ and $R = A$ or $AM = (0)$. Suppose that $AM = (0)$; then $AS \subseteq RS \subseteq R$, and either $AS = (0)$ or $AS = R$. Now $AS = (0)$ says that A is a right ideal of S and $A = R$. If $AS = R$, then $R^2 \subseteq ASAS \subseteq ASMS \subseteq AM = (0)$, which contradicts the potency of S. Hence AM cannot be (0). It follows that $R = A$; hence we have shown that a minimal right ideal of S contained in M is a minimal right ideal of M.

However, it is unlikely that the hypothesis of Theorem 2 will imply that every right ideal of M is a right ideal of S. As yet we have not found a counterexample.

If we are willing to restrict ourselves to semirings in which the lattice of right ideals is modular, then every right ideal of M is a right ideal of S. The modularity of the lattice of right ideals implies that this lattice satisfies the exchange lemma.[11]

THEOREM 3. *If S is a potent semiring in which the lattice of right ideals is modular and M is a minimal two-sided ideal containing a minimal right ideal of S, then every right ideal of M is a right ideal of S.*

Proof: We show that $xM = xS$ for $x \in M$. By Theorem 1, $M = \sum R_i$, where $R_i \in \Delta$, the set of minimal right ideals of S contained in M. Therefore, $x = x_1 + x_2 + \ldots + x_n$, n finite, and $R_i \in \Delta$. $xS \subseteq x_1S + x_2S + \ldots + x_nS \subseteq R_1 + R_2 + \ldots + R_n$. By the exchange lemma, we can renumber $R_1, R_2, \ldots, R_n$ so that $R_1 + R_2 + \ldots + R_n = R'_1 + \ldots + R_j' + R_{j+1} + \ldots + R_n$, where $R_1' + R_2' + \ldots + R_j' = xS$, holds. We replace R_i' by R_i, for $i = 1, 2, \ldots, j$, so that $xS = R_1 + R_2 + \ldots + R_j$. Since S is potent, we have $(0) \subset R_i^2 \subseteq R_i$. The minimality of R_i implies $R_i^2 = R_i$. Hence $xS \supseteq xM \supseteq xSM = (R_1 + R_2 + \ldots + R_j)M \supseteq R_1^2 + R_2^2 + \ldots + R_j^2 = R_1 + R_2 + \ldots + R_j = xS$. We have $xM = xS$.

Let R be a right ideal of M. Then $RS = RM \subseteq R$, and R is a right ideal of S.

We give an example of a potent semiring which is not modular.

Example 3: Let S_4 be the semiring of order 4 with elements $0, a, b, c$ whose addition and multiplication tables are, respectively,

$+$	0	a	b	c
0	0	a	b	c
a	a	a	a	a
b	b	a	b	b
c	c	a	b	c

$\cdot$	0	a	b	c
0	0	0	0	0
a	0	a	a	a
b	0	b	b	b
c	0	b	b	b

Then $R = \{0, a\}$, $R_1 = \{0, b, c\}$, $R_2 = \{0, b\}$ are right ideals which fail to satisfy the modular identity, i.e., $R_1 \cap (R_2 + R) \neq R_2 + R_1 \cap R$.

4. POTENT SEMIRINGS IN WHICH TWO-SIDED IDEALS CONTAIN A MINIMAL RIGHT IDEAL AND A MINIMAL LEFT IDEAL

LEMMA 2. *If M is a minimal two-sided ideal of a potent semiring S and if M contains a minimal right ideal and a minimal left ideal of S and if R is a minimal right ideal of S contained in M, then there exists a minimal left ideal L of S contained in M such that $LR = M$ and $RL \neq (0)$.*

Proof: Since the right ideal $RM \subseteq R$, $RM = R$ or $RM = (0)$. If $RM = (0)$, then $R^2 \subseteq RM = (0)$, which contradicts the potency of S. We have $RM = R$.

Since the two-sided ideal $MR \subseteq M$, $MR = M$ or $MR = (0)$. As before, the potency of S excludes $MR = (0)$. We have $MR = M$.

Theorem 1 implies that M is the sum of all minimal left ideals L of S which are contained in M. If $LR = (0)$, for all such L of S, then also $MR = (0)$, contradicting $MR = M$. Thus there exists a minimal left ideal L of S contained in M such that $LR \neq (0)$. Since the two-sided ideal $LR \subseteq M$, $LR = M$. If $RL = (0)$, then $R = RM = RLR = (0)$. Hence $RL \neq (0)$.

In the following sequence of lemmas we let *M be a minimal two-sided ideal of a potent semiring S which contains a minimal right ideal and a minimal left ideal of S, and R be a minimal right ideal and L be a minimal left ideal contained in M, such that* $LR = M$ *and* $RL \neq (0)$. *Then*

LEMMA 3. *The equations $s_1x = s_2$ and $ys_1 = s_2$ are solvable in RL, where $s_1 \neq 0$ is in $R \cap L$ and s_2 is in RL.*

Proof: Since s_1 is in L, Ms_1 is a left ideal of S contained in M. Therefore, $Ms_1 = L$ or $Ms_1 = (0)$. If $Ms_1 = (0)$, $s_1 \neq 0$, then the set X of all elements x of M for which $Mx = (0)$ contains an element $\neq 0$. Moreover, X is a two-sided ideal of S. X is evidently closed under addition. Now $(0) = Mx \supseteq (MS)x = M(Sx)$. This implies that $Sx \subseteq X$. Similarly, $M(xS) = (Mx)S = (0)$ and $xS \subseteq X$. Thus $X^2 \subseteq MX = (0)$ and $X = (0)$ and $s_1 = 0$. We have $Ms_1 = L$. Since s_1 is in R, s_1R is a right ideal of S contained in R. Therefore, $s_1R = R$ or $s_1R = (0)$. If $s_1R = (0)$, then $M = LR = Ms_1R = (0)$. We have $s_1R = R$. Finally, $s_1RL = RL$, and the equation $s_1x = s_2$ is solvable in RL. In exactly the same manner we prove that the equation $ys_1 = s_2$ is solvable in RL.

LEMMA 4. *RL is a division semiring.*

LEMMA 5. *$L = Se$, $R = eS$, and $R \cap L = eSe$, where e is the identity of the division semiring.*

LEMMA 6. *$R \cap L = RL$.*

LEMMA 7. *The identity element e of the division semiring RL is a primitive idempotent.*

The proofs of Lemmas 4–7 are straightforward and analogous to those given by Clifford,[9] and they are therefore omitted.

If each two-sided ideal $\neq (0)$ in a potent semiring contains a minimal right ideal and a minimal left ideal of S, then

LEMMA 8. *Each two-sided ideal $\neq (0)$ of S contains a minimal two-sided ideal of S (the smallest two-sided ideal containing a minimal right ideal R of S that is contoined in the given two-sided ideal).*

Proof: Let A be a two-sided ideal of S, and R a minimal right ideal of S which is contained in A. Then SR is a two-sided ideal of S contained in A. Since $R^2 = R \neq (0)$, then $SR \neq (0)$. We prove that SR is minimal. Let B be a two-sided ideal $\neq (0)$ such that $B \subseteq SR$. Then RB is a right ideal of S such that $(0) \subseteq RB \subseteq R$. Since R is a minimal right ideal of S, then either $RB = R$ or $RB = (0)$. If $RB = (0)$, then $SRB = (0)$ and $B^2 = (0)$. The potency of S requires $B = (0)$, which leads to a contradiction. Hence $R = RB \subseteq SB \subseteq B$ and $SR \subseteq SB \subseteq B$. Previously, $B \subseteq SR$. We have then $B = SR$.

LEMMA 9. *Each minimal right ideal of S is contained in a minimal two-sided ideal of S.*

Proof: Immediate consequence of the previous lemma.

LEMMA 10. *Each right ideal* $\neq(0)$ *of* S *contains a minimal right ideal of* S.

Proof: If R is a right ideal $\neq(0)$ of S, then $SR \neq (0)$, because there are no right annihilators $\neq(0)$ of S. By assumption, SR, a two-sided ideal of S, contains a minimal right ideal $R_1 \neq (0)$. As before, the set X of all elements $x \in SR$ which are right annihilators of SR is a two-sided ideal of S. Now $X^2 \subseteq SRX = (0)$. The potency of S requires that $X = (0)$. This implies that $SRR_1 \neq (0)$ and $RR_1 \neq (0)$. Therefore, there exists an element t in R such that $tR_1 \neq (0)$. By Lemma 1, tR_1 is a minimal right ideal $\neq(0)$ of S contained in R.

We are now ready to prove the main result of this paper.

THEOREM 4. *If* S *is a potent semiring in which each two-sided ideal contains a minimal right ideal and a minimal left ideal of* S, *then any right ideal* $R \neq (0)$ *contains a multiplicative idempotent.*

Proof: Lemma 10 yields a minimal right ideal R_2 in R. This minimal right ideal is contained in a minimal two-sided ideal M of S, by Lemma 9. Lemma 2 gives a minimal left ideal L of S such that $LR_2 = M$ and $R_2L \neq (0)$. R_2L is a division semiring, by Lemma 4. Its multiplicative identity $e \neq 0$ is the multiplicative idempotent $\neq 0$ contained in R. Lemma 5 states that $R_2 = eS$.

We conclude this paper with the following observations. Using Euler's argument, we can show that if S is a semiring and if R is a minimal right ideal of S consisting of finitely many elements, then either $R^2 = (0)$ or R contains an idempotent element. Hence the fundamental lemma of structure theory in rings remains true in finite semirings with minimum condition. However, this lemma does not remain true in any semiring with minimum condition. For, when we change the order of multiplication, the semigroup $\mathfrak{S}$ given by Baer and Levi[12] is an example of a semigroup with right division but containing no idempotent element. We imbed $\mathfrak{S}$ into a semiring S as follows: $S = \{0, \mathfrak{S}\}$, with s_1s_2 as in $\mathfrak{S}$ if $s_1 \neq 0$ and $s_2 \neq 0$, $s_1s_2 = 0$ otherwise, $s_1 + s_2 = s_2$ if s_2 in $\mathfrak{S}$, and $s_1 + s_2 = s_1$ if $s_2 = 0$. Then S is a potent semiring without an idempotent element and one in which there is no right ideal different from S and 0. The question whether the condition, in Theorem 4, that S be a potent semiring is necessary remains unanswered.

* Presented to the American Mathematical Society, December 28, 1954.

[1] H. S. Vandiver, "On Some Simple Types of Semi-rings," *Am. Math. Monthly*, **46**, 22–26, 1939.

[2] H. S. Vandiver, "Note on a Simple Type of Algebra in Which Cancellation Law of Addition Does Not Hold," *Bull. Am. Math. Soc.*, **40**, 920, 1934.

[3] G. Birkhoff and S. Mac Lane, *A Survey of Modern Algebra* (New York: Macmillan Co., 1941), pp. 342–344.

[4] S. Bourne, "The Jacobson Radical of a Semiring," these PROCEEDINGS, **37**, 163–170, 1951.

[5] N. Jacobson, *Theory of Rings* (Providence, R.I.: American Mathematical Society, 1943), p. 2.

[6] G. Birkhoff, *Lattice Theory* (Providence, R. I.: American Mathematical Society, 1948), p. 27.

[7] B. L. Van der Waerden, *Modern Algebra*, **2** (New York: Frederick Ungar Pub. Co., 1952), 22–23.

[8] E. Artin, C. J. Nesbitt, and R. M. Thrall, *Rings with Minimum Condition* ("University of Michigan Publications in Mathematics," Vol. 1 [Ann Arbor: University of Michigan, 1944]), pp. 15–16.

[9] A. H. Clifford, "Semigroups without Nilpotent Ideals," *Am. J. Math.*, **71**, 834–841, 1949.

[10] N. Jacobson, *Lectures in Abstract Algebra*, **1**,(New York: D. Van Nostrand Co., 1951), 189–195.

[11] S. Mac Lane, "A Lattice Form for Transcendence Degrees and P-Bases," *Duke Math. J.*, **4**, 455–468, 1938.

[12] R. Baer and F. Levi, "Vollständiger irreduzibele System von Gruppenaxiomen," *Sitzber. Heidelberg. Akad. Wiss., Math.-naturw. Kl., Beiträge Zur Algebra*, **18**, 7, 1932.

GENERALIZED RESOLUTIONS OF THE IDENTITY FOR CLOSED SYMMETRIC ORDINARY DIFFERENTIAL OPERATORS

By Earl A. Coddington*

MATHEMATICS INSTITUTE, UNIVERSITY OF COPENHAGEN, AND UNIVERSITY OF CALIFORNIA, LOS ANGELES

Communicated by Einar Hille, June 7, 1956

1. *Introduction.*—It is the purpose of this note to outline how any generalized resolution of the identity for a closed symmetric ordinary differential operator may be expressed in terms of a basis of solutions of a differential equation and a certain spectral matrix. This result implies a corresponding expansion theorem and isometry with a transform space. The proof includes the case of a self-adjoint operator, and the results reduce in this case to the known ones, except that, of course, the isometry involved in the self-adjoint case is a unitary mapping.

We consider a formal ordinary differential operator

$$L = p_0 D^n + p_1 D^{n-1} + \ldots + p_n,$$

where $D = d/dx$, the p_k are complex-valued functions of class C^{n-k} on an open real interval $a < x < b$ ($a = -\infty$ or $b = +\infty$, or both, are allowed), and $p_0(x) \neq 0$ on (a, b). We also assume that L is formally self-adjoint, i.e., L coincides with its Lagrange adjoint

$$L^+ = (-1)^n D\bar{p}_0 + (-1)^{n-1} D^{n-1}\bar{p}_1 + \ldots + \bar{p}_n.$$

Let $\mathfrak{H}$ be the Hilbert space of all complex-valued functions on (a, b) whose magnitudes are square summable on (a, b). We denote by $\mathfrak{D}$ the set of all $u \in \mathfrak{H}$ which have continuous derivatives up to order $n - 1$ on (a, b), $u^{(n-1)}$ is absolutely continuous on every closed subinterval of (a, b), and $Lu \in \mathfrak{H}$. Let T be the operator in $\mathfrak{H}$ with domain $\mathfrak{D}$, and $Tu = Lu$, $u \in \mathfrak{D}$.

For $u, v \in \mathfrak{D}$ we have Green's formula,

$$\int_y^x (\bar{v}Lu - u\overline{Lv}) = [uv](x) - [uv](y),$$

where $[uv](x)$ is a certain form in $(u, u', \ldots, u^{(n-1)})$ and $(v, v', \ldots, v^{(n-1)})$. It follows from Green's formula that $[uv](x)$ tends to a limit as x tends to a or b. These limits are denoted by $[uv](a)$ and $[uv](b)$, respectively, and we let $\langle uv \rangle = [uv](b) - [uv](a)$. Therefore, $\langle uv \rangle = (Lu, v) - (u, Lv)$, where $(\ ,\ )$ is the inner product in $\mathfrak{H}$. Let $\mathfrak{D}_0$ be the set of all $u \in \mathfrak{D}$ satisfying $\langle uv \rangle = 0$ for all $v \in \mathfrak{D}$, and define the operator T_0 by prescribing its domain as $\mathfrak{D}_0$, and $T_0 u = Lu$, $u \in \mathfrak{D}_0$. Then T_0 is closed, symmetric, its adjoint T_0^* is T, and $T^* = T_0$. In a sense, T_0 is the smallest closed symmetric operator in $\mathfrak{H}$ which is associated with L. If T_1 is any closed symmetric operator in $\mathfrak{H}$ with domain $\mathfrak{D}_1$ which contains

all functions in $\mathfrak{D}$ which vanish outside closed bounded subintervals of (a, b) and such that $T_1u = Lu$, $u \in \mathfrak{D}_1$, then $T_0 \subseteq T_1$.[1]

We consider any generalized resolution of the identity F associated with the operator T_0. A generalized resolution of the identity F is a one-parameter family $F = \{F(\lambda), -\infty < \lambda < +\infty\}$ of self-adjoint operators in $\mathfrak{H}$ satisfying (*a*) $F(\lambda) - F(\mu)$ is positive bounded if $\lambda > \mu$; (*b*) $F(-\infty) = 0$, $F(\infty) = I$ (the identity operator); and (*c*) $F(\lambda + 0) = F(\lambda)$. If the further restriction $F(\lambda)\,F(\mu) = F(\nu)$, $\nu = \min(\lambda, \mu)$, is imposed, then F becomes an ordinary resolution of the identity. The family F is associated with T_0 if for any $u \in \mathfrak{D}_0$, $v \in \mathfrak{H}$,

$$(T_0u, v) = \int_{-\infty}^{\infty} \lambda d(F(\lambda)u, v), \qquad \|T_0u\|^2 = \int_{-\infty}^{\infty} \lambda^2 d(F(\lambda)u, u).$$

The operator T_0 can always be extended to a self-adjoint operator H in a (possibly larger) Hilbert space $\mathfrak{H}_1$, and each generalized resolution of the identity F is of the form $F = PE$, where E is a resolution of the identity for an appropriate H and P is the projection of $\mathfrak{H}_1$ onto $\mathfrak{H}$. Conversely, if $F = PE$, where P and E are as above, then F is a generalized resolution of the identity for T_0.[2]

Corresponding to any generalized resolution of the identity F for T_0 there is a unique generalized resolvent $R(l)$ given by

$$(R(l)f, g) = \int_{-\infty}^{\infty} \frac{d(F(\lambda)f, g)}{\lambda - l} \qquad (f, g \in \mathfrak{H};\ \Im l \neq 0). \quad (1)$$

Using a characterization of generalized resolvents due to A. V. Štraus,[3] we indicate how the range of $R(l)$ can be prescribed by certain boundary conditions involving the parameter l. Using this, and the existence of a right inverse for $T - l$, which we obtained earlier,[4] we show that $R(l)$ is an integral operator of Carleman type, with a kernel having well-defined properties. By writing this kernel as the sum of a certain fundamental solution for $(L - l)u = 0$ and a second kernel, we then indicate that an inversion of equation (1) can be carried out to yield the expressions for the $F(\lambda)$. As indicated above, these expressions immediately imply an expansion result and Parseval equality.

To our knowledge, the only results of this type which have been obtained for F which are not ordinary resolutions of the identity are those due to A. V. Štraus[5] and ourselves.[6] A. V. Štraus considered the case $L = -DpD + q$ on $0 \leqq x < \infty$, where p and q are real and the limit-point case prevails at infinity. In this case only boundary conditions at the origin enter. He obtained the expressions for all $F(\lambda)$ in this situation. Our present method may be considered as the appropriate extension of his. In the case of a maximal symmetric extension of T_0 (which always exists in $\mathfrak{H}$) there is a unique associated generalized resolution of the identity, and we have determined[6] the expressions for the $F(\lambda)$ by a quite different method.

Detailed proofs will appear elsewhere.

2. *The Generalized Resolvent.*—Let F be an arbitrary, but fixed, generalized resolution of the identity associated with T_0, and let $R(l)$ be the corresponding generalized resolvent given by equation (1). Let $\mathcal{E}(l)$ denote the eigenspace of T for l, i.e., $\mathcal{E}(l)$ is the set of all $u \in \mathfrak{D}$ such that $Tu = lu$. If $\omega^+ = \dim \mathcal{E}(i)$, $\omega^- = \dim \mathcal{E}(-i)$, then it is known that $\dim \mathcal{E}(l) = \omega^+$ if $\Im l > 0$ and $\dim \mathcal{E}(l) = \omega^-$ if $\Im l < 0$. According to A. V. Štraus,[3] every $R(l)$ has the form

$$R(l) = (T_{A(l)} - l)^{-1} \quad (\Im l > 0), \qquad R(\bar{l}) = (R(l))^*. \tag{2}$$

Here $A(l)$ is an operator taking $\mathcal{E}(-i)$ into $\mathcal{E}(i)$, which is analytic for $\Im l > 0$ and whose norm is less than or equal to one. The operator $T_{A(l)}$ is the quasi-self-adjoint extension of T_0 generated by $A(l)$, that is, the operator whose domain $\mathfrak{D}(l)$ is the set of all $u \in \mathfrak{D}$ having the form $u = u_0 + (I - A(l))u^-$, where $u_0 \in \mathfrak{D}_0$, $u^- \in \mathcal{E}(-i)$, and I is the identity operator on $\mathcal{E}(-i)$, and which is defined by $T_{A(l)}u = Tu\ (= Lu)$ for $u \in \mathfrak{D}(l)$. Conversely, given any such $A(l)$, the operator $(T_{A(l)} - l)^{-1}$ for $\Im l > 0$ determines a generalized resolvent $R(l)$ for T_0 via equation (2).

Let $\phi_1, \ldots, \phi_{\omega^+}$ be an orthonormal basis for $\mathcal{E}(i)$, and let $\psi_1, \ldots, \psi_{\omega^-}$ be one for $\mathcal{E}(-i)$. Define the functions $v_j(l)$ and $v_j^*(l)$ by

$$\begin{aligned} v_j(l) &= \psi_j - A(l)\psi_j && (j = 1, \ldots, \omega^-), \\ v_j^*(l) &= \phi_j - (A(l))^*\phi_j && (j = 1, \ldots, \omega^+). \end{aligned}$$

THEOREM 1. *For $\Im l > 0$, the domain $\mathfrak{D}(l)$ of $T_{A(l)}$ is the set of all $u \in \mathfrak{D}$ satisfying $\langle uv_j^*(l)\rangle = 0$, $j = 1, \ldots, \omega^+$, and the domain $\mathfrak{D}^*(l)$ of $(T_{A(l)})^*$ is the set of all $u \in \mathfrak{D}$ satisfying $\langle uv_j(l)\rangle = 0$, $j = 1, \ldots, \omega^-$.*

Now consider a right inverse of $T - l$ of the type we constructed in a previous work.[4] This is an integral operator $G(l)$ of Carleman type with bound $\|G(l)\| \leqq |\Im l|^{-1}$ and a kernel $G = G(x, y, l)$ having the property that $\partial^{j+k-2}G/\partial x^{j-1}\partial y^{k-1}$ $(j, k = 1, \ldots, n)$ are continuous in (x, y, l), and analytic in l, on any region for which $\Im l \neq 0$. An exception occurs at $x = y$ when j or k is n, due to the well-defined jump there. For $f \in \mathfrak{H}$, $u = R(l)f - G(l)f \in \mathfrak{D}$ and $(T - l)u = 0$, which implies $R(l)f - G(l)f \in \mathcal{E}(l)$. Using the additional fact that $\langle R(l)fv_j^*(l)\rangle = 0$, $j = 1, \ldots, \omega^+$, we find that $R(l)f$ may be written as

$$R(l)f = G(l)f + \sum_{j=1}^{\omega^-}\sum_{k=1}^{\omega^+} \Phi_{kj}(l)(f, \psi_j(\bar{l}))\,\phi_k(l) \qquad (\Im l > 0), \tag{3}$$

where $\phi_1(l), \ldots, \phi_{\omega^+}(l)$ and $\psi_1(l), \ldots, \psi_{\omega^-}(\bar{l})$ are bases for $\mathcal{E}(l)$ and $\mathcal{E}(\bar{l})$, respectively. From this representation we can deduce the following result.

THEOREM 2. *The generalized resolvent $R(l)$ is an integral operator of Carleman type having a kernel $K = K(x, y, l)$ with the property that $\partial^{j+k-2}K/\partial x^{j-1}\partial y^{k-1}$ $(j, k = 1, \ldots, n)$ are continuous in (x, y, l), and analytic in l, on any region for which $\Im l \neq 0$, except for $x = y$, when j or k is n.*

The analytic nature of the kernel follows by introducing analytic bases for $\mathcal{E}(l)$ and $\mathcal{E}(\bar{l})$. If $\Im l_0 > 0$, then $\phi_k(l)$, $\psi_j(\bar{l})$ defined by

$$\begin{aligned} \phi_k(l) &= [I + (l - l_0)\,G(l)]\,\phi_k(l_0) && (k = 1, \ldots, \omega^+), \\ \psi_j(\bar{l}) &= [I + (\bar{l} - \bar{l}_0)\,G(\bar{l})]\psi_j(\bar{l}_0) && (j = 1, \ldots, \omega^-), \end{aligned}$$

will be analytic bases for $\mathcal{E}(l)$, $\mathcal{E}(\bar{l})$, respectively, if $|l - l_0| < |\Im l_0|/2$. Here $\phi_k(l_0)$, $\psi_j(\bar{l}_0)$ represent any basis for $\mathcal{E}(l_0)$, $\mathcal{E}(\bar{l}_0)$, respectively. Using these bases in equation (3), it can be shown that Φ_{jk} is analytic for $|l - l_0| < |\Im l_0|/2$.

We now rewrite the kernel K in a different way, by letting $K = K_0 + K_1$, where K_0 is a certain fundamental solution for $(L - l)u = 0$, which we describe. Let $s_1(l), \ldots, s_n(l)$ be a basis for the solutions of $(L - l)u = 0$ satisfying $s_j^{(k-1)}(c, l) = \delta_{jk}$ for some c, $a < c < b$. Then the $s_j^{(k-1)}(x, l)$ are entire in l for any fixed x in (a, b). Also, $[s_j(l)\,s_k(\bar{l})](x)$ is independent of x and l and will be denoted by S_{kj}.

The matrix $S = (S_{jk})$ is nonsingular; let $S^{-1} = (S_{jk}^{-1})$ denote its inverse. We define K_0 by

$$K_0(x, y, l) = \overline{K_0(y, x, \bar{l})} = {}^1/_2 \sum_{j,k=1}^{n} S_{kj}^{-1} s_k(x, l) \overline{s_j(y, \bar{l})} \quad (y \geqq x).$$

Using this, it follows that K_1 can be written as

$$K_1(x, y, l) = \sum_{j,k=1}^{n} \Psi_{kj}(l) s_k(x, l) \overline{s_j(y, \bar{l})},$$

and the following theorem obtains.

THEOREM 3. *The matrix $\Psi = (\Psi_{jk})$ is analytic for $\Im l \neq 0$, $\Psi^*(l) = \Psi(\bar{l})$, and $\Im\Psi(l)/\Im l \geqq 0$, where $\Im\Psi = (\Psi - \Psi^*)/2i$.*

The proof results from noting that

$$\Psi_{jk}(l) = \frac{\partial^{j+k-2} K_1}{\partial x^{j-1} \partial y^{k-1}}(c, c, l)$$

and

$$(\Im\Psi)_{jk}(l) = \frac{\partial^{j+k-2} H}{\partial x^{j-1} \partial y^{k-1}}(c, c, l), \tag{4}$$

where $H(x, y, l) = (K(x, y, l) - K(x, y, \bar{l}))/2i$. The positivity of the matrix on the right-hand side of equation (4) for $\Im l > 0$ follows by applying the inequality $(\Im(R(l)f, f)/\Im l) \geqq \|R(l)f\|^2$ to suitable functions f.

3. *The Generalized Resolution of the Identity.*—From Theorem 3 we can assert that $\Psi(l)$ has the representation

$$\Psi(l) = \alpha + \beta l + \int_{-\infty}^{\infty} \frac{\lambda l + 1}{\lambda - l} \, d\sigma(\lambda) \qquad (\Im l \neq 0), \tag{5}$$

where α, β are Hermitian constant matrices, $\beta \geqq 0$, and σ is a Hermitian matrix function of bounded variation which is nondecreasing, in the sense that $(\sigma\xi, \xi)$ is monotone nondecreasing for any complex n-dimensional vector ξ. This assertion follows by applying the known theorem[7] for scalar functions to the form $(\Psi\xi, \xi)$.

From the representation (5) the limit

$$\rho(\lambda) = \lim_{\epsilon \to +0} \frac{1}{\pi} \int_0^{\lambda} \Im\Psi(\nu + i\epsilon) \, d\nu$$

exists for any fixed real λ, and the matrix ρ is nondecreasing and of bounded variation on every finite λ-interval. Using the representation $K = K_0 + K_1$ and the inversion of equation (1)

$$(F(\Delta)f, g) = \lim_{\epsilon \to +0} \frac{1}{\pi} \int_{\Delta} \Im(R(\nu + i\epsilon)f, g) \, d\nu,$$

where $\Delta = (\mu, \lambda]$ and λ, μ are continuity points of F, we arrive at the following representation for F.

THEOREM 4. *If $f \in \mathfrak{C}$ and vanishes outside a closed bounded subinterval of (a, b) and λ, μ are continuity points of ρ and F, then*

$$F(\Delta)f(x) = \int_\Delta \sum_{j,k=1}^{n} s_j(x, \lambda)\, \hat{f}_k(\lambda)\, d\rho_{jk}(\lambda),$$

where $\hat{f}_k(\lambda) = (f, s_k(\lambda))$.

Let $\zeta = (\zeta_1, \ldots, \zeta_n)$, $\eta = (\eta_1, \ldots, \eta_n)$ be vector functions of λ, and introduce the inner product

$$(\zeta, \eta) = \int_{-\infty}^{\infty} \sum_{j,k=1}^{n} \overline{\eta_j(\lambda)}\, \zeta_k(\lambda)\, d\rho_{jk}(\lambda)$$

and norm $\|\zeta\| = (\zeta, \zeta)^{1/2}$. Let $L^2(\rho)$ be the Hilbert space of all ζ, measurable with respect to ρ, such that $\|\zeta\| < \infty$. Then an immediate consequence of Theorem 4 is the following expansion result and Parseval equality.

THEOREM 5. *If $f \in \mathfrak{H}$, the vector $\hat{f} = (\hat{f}_j)$ converges in norm in $L^2(\rho)$ and $\|f\| = \|\hat{f}\|$. In terms of this $\hat{f}$,*

$$f(x) = \int_{-\infty}^{\infty} \sum_{j,k=1}^{n} s_j(x, \lambda)\, \hat{f}_k(\lambda)\, d\rho_{jk}(\lambda),$$

where the integral converges in norm in $\mathfrak{H}$.

* The author is a Fulbright grantee.

[1] For the facts concerning T_0 and T see E. A. Coddington, "The Spectral Representation of Ordinary Self-adjoint Differential Operators," *Ann. Math.*, **60**, 192, 1954.

[2] For a discussion of generalized resolutions of the identity and generalized resolvents see, e.g., N. I. Achieser and I. M. Glasmann, *Theorie der linearen Operatoren im Hilbert-Raum* (Berlin, 1954) (German translation).

[3] A. V. Štraus, "Generalized Resolvents of Symmetric Operators," *Izvest. Akad. Nauk S.S.S.R.*, Ser. Mat., **18**, 51, 1954.

[4] E. A. Coddington, "The Spectral Matrix and Green's Function for Singular Self-adjoint Boundary Value Problems," *Can. J. Math.*, **6**, 169, 1954.

[5] A. V. Štraus, "On Spectral Functions of Differential Operators," *Izvest. Akad. Nauk S.S.S.R.*, Ser. Mat., **19**, 201, 1955.

[6] E. A. Coddington, "On Maximal Symmetric Ordinary Differential Operators," *Math. Scand.* (*to appear*).

[7] See, e.g., M. H. Stone, *Linear Transformations in Hilbert Space* (New York, 1932), Theorem 10.36, p. 573.

---

*CAUCHY'S PROBLEM FOR LINEAR DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS**

BY LEON EHRENPREIS

INSTITUTE FOR ADVANCED STUDY

Communicated by Deane Montgomery, June 27, 1956

Global solutions for Cauchy's problem in linear partial differential equations have been essentially restricted to the hyperbolic case. In the present note, we wish to show what the well-posed Cauchy problem is for linear equations with constant coefficients which may or may not be hyperbolic and which may even be parabolic or irregular (see below). As a by-product of our investigations, we shall give a new proof of Gårding's theorem that Cauchy's problem is well posed in the C^∞

theory if and only if the differential operator is hyperbolic.[1] Also, our methods are very amenable to explicit calculation of Green's functions and the solution of Cauchy's problem.

In this note we shall restrict ourselves to the case where the initial hypersurface is a hyperplane. In forthcoming papers we shall extend the results of this note in many directions, including the replacing of Cauchy's problem by more general types of boundary-value problems and the giving of boundary-value data to systems of equations which may even be degenerate.[2]

Let us denote by R (C) real (complex) Euclidean space of dimension $n + 1$. We denote the co-ordinates in R (C) by $t, x_1, x_2, \ldots, x_n$. We also denote by R' (C') a real (complex) Euclidean space with co-ordinates $s, z_1, z_2, \ldots, z_n$. We shall sometimes write x for $(x_1, \ldots, x_n)$ and z for $(z_1, \ldots, z_n)$.

By $\mathcal{E}$ we denote the space of C^∞ functions on R; $\mathcal{E}_0$ is the subspace of $\mathcal{E}$ consisting of those functions which are independent of t. We topologize $\mathcal{E}$ and $\mathcal{E}_0$ according to the method of L. Schwartz. $\mathcal{E}'$ is the dual of $\mathcal{E}$, $\boldsymbol{E}'$ is the Fourier transform[3] of $\mathcal{E}'$, and $\boldsymbol{E}$ is the dual of $\boldsymbol{E}'$. $\boldsymbol{E}_0'$ is the Fourier transform of the dual $\mathcal{E}_0'$ of $\mathcal{E}_0$, and $\boldsymbol{E}_0$ is the dual of $\boldsymbol{E}_0'$. We consider $\boldsymbol{E}_0'$ and $\boldsymbol{E}'$ as spaces of functions on C'; $\boldsymbol{E}'$ consists of all entire functions of exponential type which are of polynomial growth on R'. The Fourier transform is defined in such a manner that $\boldsymbol{E}_0'$ consists of those functions of $\boldsymbol{E}'$ which are independent of s.

By $\mathcal{C}$ we denote the space of entire functions on C with its usual topology,[3] $\mathcal{C}'$ is the dual of $\mathcal{C}$, $\boldsymbol{H}'$ is the Fourier transform of $\mathcal{C}'$, and $\boldsymbol{H}$ is the dual of $\boldsymbol{H}'$. The spaces $\mathcal{C}_0$, $\mathcal{C}_0'$, $\boldsymbol{H}_0'$, $\boldsymbol{H}_0$ bear the same relation to $\mathcal{C}$, $\mathcal{C}'$, $\boldsymbol{H}'$, $\boldsymbol{H}$ as the spaces $\mathcal{E}_0$, $\mathcal{E}_0'$, $\boldsymbol{E}_0'$, $\boldsymbol{E}_0$ do to the respective spaces $\mathcal{E}$, $\mathcal{E}'$, $\boldsymbol{E}'$, $\boldsymbol{E}$.

Now let D be a partial differential operator with constant coefficients. We assume first that D is *regular* (with respect to t); that is, we may write $D = \partial^m/\partial t^m + D_0(\partial^{m-1}/\partial t^{m-1}) + \ldots + D_m$, where the D_j are partial differential operators which are independent of t. *Cauchy's problem* then consists of finding a function f of (t, x) which satisfies

$$Df = g, \tag{1}$$

$$\frac{\partial^k f(x)}{\partial t^k} = g_k(x) \qquad \text{for } k = 0, 1, \ldots, m - 1, \tag{2}$$

where g is a given function of (t, x) and the g_k are given functions of x.

1. *Nonparabolic Equations.*—We assume that $g \in \mathcal{C}$ and that the g_k are in $\mathcal{C}_0$; we seek a solution $f \in \mathcal{C}$. We write D' for the adjoint of D. Then equation (1) states that $D'W \cdot f = W \cdot Df = W \cdot g$ for any $W \in \mathcal{C}'$. That is, on identifying $\mathcal{C}$ with the dual of $\mathcal{C}'$ (which is permissible, since $\mathcal{C}$ is reflexive), f is prescribed on the subspace $D'\mathcal{C}'$ of $\mathcal{C}'$, namely,

$$f \cdot D'W = g \cdot W. \tag{3}$$

Equation (2) tells us that, if $T \in \mathcal{C}'$ is of the form $T = T_0 + (\partial/\partial t)T_1 + \ldots + (\partial^{m-1}/\partial t^{m-1})T_{m-1}$, where $T_j \in \mathcal{C}_0'$, then the value $f \cdot T$ is given by

$$f \cdot T = \sum g_k \cdot T_k. \tag{4}$$

Let us take the Fourier transform of conditions (3) and (4). We denote by P

the Fourier transform of D; hence P is a polynomial which we may write as $P(s, z) = s^m + P_1(z)s^{m-1} + \ldots + P_m(z)$, where the P_j are polynomials. We are looking for a $U \in \boldsymbol{H}$ which satisfies

$$U \cdot PF = V \cdot F \tag{5}$$

for all $F \in \boldsymbol{H}'$, where $V \in \boldsymbol{H}$ is given, and, if $G \in \boldsymbol{H}'$ is of the form $G(s, z) = G_0(z) + G_1(z)s + \ldots + G_{m-1}(z)s^{m-1}$, where $G_k \in \boldsymbol{H}_0'$, then

$$U \cdot G = \sum V_k \cdot G_k, \tag{6}$$

where the V_k are prescribed elements of $\boldsymbol{H}_0$.

The solution of our problem would be complete if we could show that every $K \in \boldsymbol{H}'$ admits a unique decomposition in the form

$$K = \sum_{j=1}^{m-1} G_j s^j + PF, \tag{7}$$

where the $G_j \in \boldsymbol{H}_0'$ and $F \in \boldsymbol{H}'$ depend continuously on K; if this is the case, we say that D is *nonparabolic* (with respect to t). The decomposition (7) is the analogue for the space $\boldsymbol{H}'$ of the Euclidean algorithm.

A simple argument shows that any $K \in \boldsymbol{H}'$ admits a decomposition in the form (7), where the G_j and F are entire functions on C' which are uniquely determined by K; we want to know when G_j and F lie in $\boldsymbol{H}'$. To answer this, we note that the G_j are determined by the values of K on the algebraic manifold M defined by the equation $P = 0$ if P has no multiple polynomial factors, or else by the values of K on M together with certain derivatives of K on parts of M. We can easily see that there exists a polynomial $P^0(z)$ (the discriminant of P) such that, for each j, P^0G_j can be expressed in the form

$$P^0(z)\, G_j(z) = \sum_{r=1}^{m} K(s_r(z))Q_j(s_r(z)) \tag{8}$$

if P has no multiple factors (with a similar expression if P has multiple factors, except that some of the derivatives of K appear on the right-hand side of eq. 8). Here the Q_j are polynomials in one variable, and the $s_k(z)$ are the zeros of the equation $P(s, z) = 0$. Now $\boldsymbol{H}'$ is the space of all entire functions of exponential type. From equation (8) it follows that P^0G_j will be in $\boldsymbol{H}_0'$ if, for all r, $s_r(z) = O(|z|)$. From $P^0G_j \in \boldsymbol{H}_0'$, we deduce[3] also that $G_j \in \boldsymbol{H}_0'$, and, from equation (7), also that $F \in \boldsymbol{H}'$. The same argument also shows that the G_j and F depend continuously on K in this case.

Conversely, if we do not have $s_r(z) = O(|z|)$ for all r, then we can construct a $K \in \boldsymbol{H}'$ such that, for some l, $G_l \notin \boldsymbol{H}_0'$.

The condition $s_r(z) = O(|z|)$ for all r is easily seen to be equivalent to the condition that m be the degree of the polynomial P. Thus we have

THEOREM 1. *D is nonparabolic if and only if m is equal to the order of the differential operator D.*

Whether D is parabolic or not, the set of K of the form (7) with G_j and F in $\boldsymbol{H}'$ contains all polynomials and so is dense in $\boldsymbol{H}'$. It follows from topological arguments that we can find g and g_j such that Cauchy's problem is not solvable if D is parabolic.

THEOREM 2. *If D is nonparabolic, then Cauchy's problem is well posed in $\mathcal{K}$, that is, for any $g \in \mathcal{K}$, $g_j \in \mathcal{K}_0$, there exists a unique solution in $\mathcal{K}$ to equations* (1) *and* (2) *which, moreover, depends continuously on g and the g_j. In the case that D is parabolic, we can find a $g \in \mathcal{K}$ and $g_j \in \mathcal{K}_0$ such that equations* (1) *and* (2) *have no solution in $\mathcal{K}$. In any case, equations* (1) *and* (2) *have at most one solution in $\mathcal{K}$ for g_j, $g \in \mathcal{K}$.*

2. *Hyperbolic Equations.*—We are looking for those operators D for which we have an analogue of Theorem 2 above for $\mathcal{E}$ in place of $\mathcal{K}$. D is called *hyperbolic* (with respect to t) if, for each $K \in \boldsymbol{E}'$, we have an expansion (7), where the G_j and F lie in $\boldsymbol{E}'$ and where the G' and F depend continuously on K. We repeat the same type of argument as above for $\mathcal{E}$ instead of $\mathcal{K}$; we want to find those operators D for which, as in equation (8) above,

$$P^0(z)G_j(z) = \sum_{r=1}^{m} K(s_r(z))\, Q_j(s_r(z)) \tag{8'}$$

(with suitable modifications if P has multiple factors) lies in $\boldsymbol{E}_0'$ whenever $K \in \boldsymbol{E}'$. It follows from the definition of the space $\boldsymbol{E}'$ that this will be the case if, for all r, $\Im(s_r(z)) = O(1 + |\Im(z)|)$, where $\Im$ denotes the imaginary part. Again,[3] if $P^0G_j \in \boldsymbol{E}'$, then $G_j \in \boldsymbol{E}'$ and hence also $F \in \boldsymbol{E}'$.

On the other hand, if, for some r, $\Im(s_r(z)) \neq O(1 + |\Im(z)|)$, then we can construct a $K \in \boldsymbol{E}'$ and an l such that $G_l \notin \boldsymbol{E}'$.

THEOREM 3. *D is hyperbolic if and only if $\Im(s_r(z)) = O(1 + |\Im(z)|)$ for all r.*

Contrary to the case of the space $\boldsymbol{H}'$ above, the set of functions $K \in \boldsymbol{E}'$ which have an expansion (7) with G_j, $F \in \boldsymbol{E}'$, may not be dense in $\boldsymbol{E}'$. This set will be dense if and only if the hypersurface $t = 0$ is not a characteristic[4] of D.

THEOREM 4. *If D is hyperbolic, then the Cauchy problem is well posed for E. If D is not hyperbolic, then either $t = 0$ is a characteristic of D, or we can find $g \in \mathcal{E}$ and $g_j \in \mathcal{E}_0$ such that equations* (1) *and* (2) *have no solution in $\mathcal{E}$.*

Remark: It is not difficult to show that every hyperbolic operator is nonparabolic.

3. *Nonparabolic, Nonhyperbolic Operators.*—Let $0 < \alpha \leqq 1$ be given; we choose the space $\mathcal{S}_\alpha$ which lies between H and E; $\mathcal{S}_\alpha$ (which is a space of non-quasi-analytic functions) is so defined that the Fourier transform $\boldsymbol{G}_\alpha'$ of its dual consists of all entire functions of exponential type F which satisfy $F(s, z) = O(e^{|s|^\alpha + |z|^\alpha})\, Q(s, z)$ for $(s, z) \in R'$, for some polynomial Q. Then, by an analysis like that above, we can find those operators for which we have a well-posed Cauchy problem for the space $\mathcal{S}_\alpha$.

4. *Parabolic Operators.*—In order to get a well-posed Cauchy problem, we have to consider spaces for which the Fourier transform of the dual is no longer a space of entire functions of exponential type. Examples are given by the spaces Q_A of entire functions of order $\leqq A$ and by many of the spaces which arise in the theory of infinite derivatives.[3] The important point to note is that, in the case of nonparabolic operators, the Cauchy data g, g_k are restricted only from the point of regularity, while in the parabolic case the g and g_k are restricted in regularity and growth at infinity. This theory is, of course, well known for the heat equation.

5. *Irregular Operators.*—If D is irregular, i.e., $D = D_0(\partial^m/\partial t^m) + D_1(\partial^{m-1}/\partial t^{m-1}) + \ldots + D_m$, where the D_k are linear differential operators in x with constant coefficients, $D_0 \neq 0, 1$, so that $P = P_0 s^m + \ldots + P_m$, where the P_k are polynomials

in z with $P_0 \neq 0, 1$, then we cannot hope to get an expansion (7) in general. What has to be done is to add to the right-hand side of equation (7) some terms which depend on the polynomial P_0. Thus the "initial surface" for the Cauchy data gets replaced by several surfaces. The method then proceeds as before.

* Work supported by National Science Foundation Grant NSF 5-G1010 at the Institute for Advanced Study.

[1] "Linear Hyperbolic Equations with Constant Coefficients," *Acta Math.*, **85**, 1–62, 1951. Actually, our result (Theorem 4) is more powerful than Gårding's because he had to assume that the Cauchy problem is well posed for ε in order to deduce that D is hypberbolic.

[2] L. Ehrenpreis, "Solutions of Some Problems of Division. I," *Am. J. Math.*, **76**, 883–903, 1954.

[3] L. Ehrenpreis, "Solution of Some Problems of Division. III," *ibid.* (to appear), and "Theory of Infinite Derivatives" (to appear).

[4] L. Hörmander, "General Partial Differential Operators," *Acta Math.*, **94**, 161–248, 1955. See especially pp. 216–217.

A UNIFIED THEORY OF GENERAL RELATIVITY OF GRAVITATION AND ELECTROMAGNETISM. II

By Luther P. Eisenhart

FINE HALL, PRINCETON UNIVERSITY

Communicated June 28, 1956

This is another proposal of a unified theory.[1] The theory is based upon a symmetric tensor g_{ij}, a coefficient of connection Γ^i_{jk}, and an electromagnetic vector φ_i, in terms of which there is the skew-symmetric electromagnetic field tensor

$$F_{ij} = \varphi_{i,j} - \varphi_{j,i} = \varphi_{i;j} - \varphi_{j;i}. \tag{1}$$

Here, and throughout the paper, a comma followed by an index denotes the derivative with respect to x with this index, and a semicolon followed by an index denotes the corresponding covariant derivative in terms of the Christoffel symbols in g_{ij}.

From equation (1) it follows that

$$F_{ij,k} + F_{jk,i} + F_{ki,j} = 0, \tag{2}$$

from which one has

$$F_{ij;k} + F_{jk;i} + F_{ki;j} = 0. \tag{3}$$

It being required that the determinant g of g_{ij} be different from zero, and in fact negative, a contravariant symmetric tensor g^{ij} is defined by $g^{ij}g_{ik} = \delta^i_k$ where δ^i_k is the Kronecker delta.

For any coefficients of connection Γ^i_{jh} one has

$$\Gamma^i_{jh} = \left\{ {i \atop jh} \right\} + a_{jh}{}^i, \tag{4}$$

where $a_{jh}{}^i$ is a tensor, as follows from the equation for the Christoffel symbols $\left\{ {i \atop jh} \right\}$ and Γ^i_{jh} separately in two co-ordinate systems.[2]

For coefficients of connection Γ^i_{jh} there is the generalized tensor[3]

$$\Gamma^i_{jhl} = \Gamma^i_{jl,h} - \Gamma^i_{jh,l} + \Gamma^k_{jl}\Gamma^i_{kh} - \Gamma^k_{jh}\Gamma^i_{kl}. \tag{5}$$

For Γ^i_{jh} defined by equation (4), this is

$$\Gamma^i_{jhl} = R^i_{jhl} + a_{jl}{}^i{}_{;h} - a_{jh}{}^i{}_{;l} + a_{jl}{}^k a_{kh}{}^i - a_{jh}{}^k a_{kl}{}^i, \tag{6}$$

where R^i_{jhl} is the Riemann tensor. Contracting equation (6) for i and l, one has

$$\Gamma_{jh} = \Gamma^i_{jhi} = R_{jh} + a_{ji}{}^i{}_{;h} - a_{jh}{}^i{}_{;i} + a_{ji}{}^k a_{kh}{}^i - a_{jh}{}^k a_{ki}{}^i, \tag{7}$$

where R_{jh} is the Ricci tensor.

Thus far the treatment is the same as in the former paper.[1] The new proposals follow.

We require that

$$\Gamma_{\underline{jh}} = 0, \tag{8}$$

where $\Gamma_{\underline{jh}}$ denotes the symmetric part of Γ_{jh}. From (7) this is

$$R_{jh} = -{}^1/_2\,(a_{ji}{}^i{}_{;h} + a_{hi}{}^i{}_{;j} + a_{ji}{}^k a_{kh}{}^i + a_{hi}{}^k a_{kj}{}^i) + {}^1/_2\,(a_{jh}{}^i{}_{;i} + a_{hj}{}^i{}_{;i}) + {}^1/_2(a_{jh}{}^k + a_{hj}{}^k)a_{ki}{}^i. \tag{9}$$

We take

$$a_{jh}{}^i = F_{jh}\lambda^i, \tag{10}$$

where for the present the vector λ^i is not specified. In terms of quantities of the type (10), equation (9) is

$$R_{jh} = {}^1/_2\,[F^i{}_j(\lambda_{i;h} + \lambda_i\lambda_k F^k{}_h) + F^i{}_h(\lambda_{i;j} + \lambda_i\lambda_k F^k{}_j)] + {}^1/_2\,\lambda^i(F_{ij;h} + F_{ih;j}). \tag{11}$$

We take

$$\lambda_{i;h} = \lambda_i\lambda_k F_h{}^k. \tag{12}$$

Equation (11) is now

$$R_{jh} = {}^1/_2\lambda^i(F_{ij;h} + F_{ih;j}). \tag{13}$$

From equation (12) one has

$$(\lambda^i\lambda_i)_{,h} = 2\lambda^i\lambda_i\lambda_k F_h{}^k,$$

which is satisfied when we take

$$\lambda^i\lambda_i = 0. \tag{14}$$

From equation (12) one has

$$\lambda_{i;hl} - \lambda_{i;lh} = \lambda_i\lambda^k(F_{hk;l} - F_{lk;h}) = \lambda_i\lambda^k F_{hl;k} \tag{15}$$

by equation (3). By means of the equation[4]

$$\lambda_{i;hl} - \lambda_{i;lh} = \lambda^j R_{jihl}$$

equation (15) becomes

$$\lambda^j R_{jihl} = \lambda_i\lambda^k F_{hl;k}.$$

Multiplying by g^{il} and contracting for i and l, one has, on changing dummy indices

$$-\lambda^j R_{jh} = \lambda^i \lambda^j F_{hi;j}.$$

From equation (13) one has

$$\lambda^j R_{jh} = {}^1/_2 \lambda^i \lambda^j F_{ih;j}.$$

Adding these equations, one obtains

$${}^1/_2 \lambda^i \lambda^j F_{hi;j} = 0. \tag{16}$$

This equation is satisfied, in consequence of equation (14), when we take

$$\lambda^i F_{ih;j} = a\lambda_h \lambda_j, \tag{17}$$

where a is a constant,

From equation (12) it follows that

$$\lambda^i{}_{;i} = 0. \tag{18}$$

By means of equation (17), equation (13) becomes

$$R_{jh} = a\lambda_j \lambda_h, \tag{19}$$

and in consequence of equation (14)

$$R = g^{jh} R_{jh} = 0. \tag{20}$$

From equation (19) and equation (12) one has

$$\begin{aligned} R^j{}_{h;j} &= a(\lambda^j \lambda_l)\,;j \\ &= a(\lambda_h \lambda^j{}_{;j} + \lambda^j \lambda_h \lambda_k F_j{}^k) = 0. \end{aligned}$$

in consequence of equation (18).

Thus R_{jh}, given by equation (19) in terms of the vector λ_i defined by equation (12), satisfies the fundamental equation[5]

$$R^j{}_{h;j} - {}^1/_2 R_{,h} = 0. \tag{21}$$

As another approach to the problem, we take

$$\lambda_{i;\,h} = \lambda_i \lambda^k F_{hk} + F_{ih} \tag{22}$$

in place of equations (12). If we put

$$\lambda^i \lambda_i = a,$$

we have from equations (22)

$$a,\, h = 2(a - 1)\lambda^k F_{hk},$$

which is satisfied by $a = 1$, and accordingly we take

$$\lambda^i \lambda_i = 1. \tag{23}$$

From equations (22) one has

$$\lambda_{i;\,hl} - \lambda_{i;\,lh} = \lambda^k (F_{hk} F_{il} - F_{lk} F_{ih}) + \lambda_i \lambda^k (F_{hk;\,l} - F_{lk;\,h}) + F_{ih;\,l} - F_{il;\,h},$$

which, in consequence of the identities (3) and the equation.[4]

$$\lambda_{i;\,hl} - \lambda_{i;\,lh} = \lambda^j R_{jihl},$$

is

$$\lambda^j R_{jihl} = \lambda^k (F_{hk}F_{il} - F_{lk}F_{ih}) + \lambda_i \lambda^k F_{hl;\,k} + F_{lh;\,i}.$$

When this equation is multiplied by g^{il} and contracted for i and l, one has

$$-\lambda^j R_{jh} = \lambda^k F_{ki} F^i{}_h + \lambda^l \lambda^k F_{hl;\,k} + F^i{}_{h;\,i}. \tag{24}$$

By means of equations (22), equation (11) is

$$R_{jh} = F_{ji}F_h{}^i + {}^1/_2 \lambda^i (F_{ij;\,h} + F_{ih;\,j}). \tag{25}$$

When this equation is multipled by λ^j and contracted for j, one has

$$\lambda^j R_{jh} = \lambda^j F_{ji} F_h{}^i + {}^1/_2 \lambda^l \lambda^j F_{ih;\,j}.$$

When this equation is added to equation (24), the result is

$${}^1/_2 \lambda^l \lambda^k F_{hl;\,k} + F^i{}_{h;\,i} = 0. \tag{26}$$

We take

$$F_{lh;\,k} = \lambda_l \lambda_k \theta_h - \lambda_h \lambda_k \theta_l, \tag{27}$$

where θ_h is a covariant vector defined later. These equations satisfy the identities (3).

By means of equations (27) and (23), equation (26) becomes

$${}^1/_2 (\theta_h - \lambda_h \lambda^l \theta_l) = F^i{}_{h;\,i}.$$

From equation (27) one has, in consequence of equation (23),

$$F^i{}_{h;\,i} = \theta_h - \lambda_h \lambda^l \theta_l.$$

From these two equations it follows that

$$F^i{}_{h;\,i} = 0 \tag{28}$$

and

$$\theta_h - \lambda_h \lambda^l \theta_l = 0. \tag{29}$$

From equation (27) by means of equation (23) one has

$$\lambda^l F_{lh;\,j} = \lambda_j (\theta_h - \lambda_h \lambda^l \theta_l) = 0.$$

In consequence of this result, equation (25) becomes

$$R_{jh} = F_{ji} F_h{}^i, \tag{30}$$

and one has

$$R = g^{jh} R_{jh} = F^{ji} F_{ji}. \tag{31}$$

From equation (30) one has

$$R^j{}_{h;\,j} = F^j{}_{i;\,j} F_h{}^i + F^{ji} F_{hi;\,j} = F^{ji} F_{hi;\,j}$$

in consequence of equation (28). From equation (31) one has

$$R_{,\,h} = 2F_{jt;\,h}F^{jt}.$$

Accordingly the left-hand member of equation (21) is

$$F^{jt}(F_{ht;\,j} - F_{jt;\,h}) = F^{jt}F_{hj;\,t}$$

in consequence of equation (3). By means of equations (27) and (29),

$$\begin{aligned} F^{jt}F_{hj;\,t} &= F^{jt}(\lambda_h\lambda_t\theta_j - \lambda_j\lambda_t\theta_h) \\ &= F^{jt}\lambda_h\lambda_t\theta_j = F^{jt}\lambda_h\lambda_t\lambda_j\lambda^l\theta_l = 0. \end{aligned}$$

Hence equation (21) is satisfied,

When equation (29) is written

$$\theta_t(\delta_h{}^l - \lambda_h\lambda^l) = 0,$$

where $\delta_h{}^l$ is the Kronecker delta, determination of the vector θ_i in terms of the vector, λi, is readily found, noting that in consequence of equation (23) the determinant of these equations is equal to zero.

[1] Cf. these PROCEEDINGS, **42**, 249–251, 1956.

[2] L. P. Eisenhart, "Non-Riemannian Geometry," *Am. Math. Soc. Pubs.*, No. 8, pp. 2, 3.

[3] *Ibid.*, p. 5.

[4] L. P. Eisenhart, *Riemannian Geometry* (Princeton, N.J.: Princeton University Press, 1926), p. 30.

[5] *Ibid.*, p. 82.

---

*SINGULAR INTEGRAL EQUATIONS FOR DIFFERENTIAL FORMS ON RIEMANNIAN MANIFOLDS**

BY J. J. KOHN

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY

Communicated by N. E. Steenrod, June 11, 1956

In the present note we obtain a necessary and sufficient condition for the regularizability of singular integral equations for differential forms on a compact Riemannian manifold. This result is applied, in a forthcoming paper by D. C. Spencer and the author, to a class of boundary-value problems for complex operators.

Notation.—Throughout this note the following notation will be used: M is a compact C^∞ Riemannian manifold of dimension m, T_x is the tangent space at $x \in M$, the inner product and the norm on T_x are denoted by $u \cdot v$ and $|u|$, respectively, S is the tangent sphere bundle over M with fiber $S_x \subset T_x$ consisting of the unit vectors in T_x, A^p is the bundle of complex-valued covariant skew-symmetric tensors of degree p, and $r(x, y)$ is the geodesic distance between x and y in M. Finally, if N is a C^∞ manifold and B is a bundle over N, then $\Gamma(B, N)$ denotes the C^∞ sections of B over N.

Consider the projection maps $\pi_1, \pi_2 : M \times M \to M$; let $A_1{}^p, A_2{}^p$ be the induced bundles over $M \times M$ defined by $\pi_1{}^{-1}A^p$, $\pi_2{}^{-1}A^p$, respectively.[1] We identify the restriction of $A_1{}^p \otimes A_2{}^p$ to the diagonal of $M \times M$ with $A^p \otimes A^p$.

Definition 1: An operator J on differential forms of degree p is called a *singular operator* if it can be expressed by

$$J\varphi(x) = \xi(x)\cdot\varphi(x) + \int_M G(x, y)\wedge * \varphi(y),$$

where $\varphi \in \Gamma(A^p, M)$; $\xi \in \Gamma(A^p \otimes A^p, M)$; $G \in \Gamma(A_1{}^p \otimes A_2{}^p, M \times M - \{\text{diagonal}\})$; $\xi(x)\cdot\varphi(x)$ means the contraction of φ with the last p indices of ξ, that is, if $\xi = \sum_k \alpha_x \otimes \beta_x$, then $\xi\cdot\varphi = \sum_k (\beta_x\cdot\varphi)\alpha_k$; the integral is taken in the sense of a principal value; and G satisfies the following conditions:

a) For each $u \in T_x - \{0\}$, $\lim_{\substack{u \\ x\to y}} r^m(x, y)\, G(x, y) = F_x(u)$ (lim denotes the limit as y approaches x along the direction of u); if $c > 0$, then $F_x(cu) = F_x(u)$.

b) The map $F: S \to A^p \otimes A^p$ defined by $F|S_x = F_x$ is a C^∞ bundle map.

c) $\int_{S_x} F_x = 0$.

d) There exists $h>0$ such that if $x, y \in M$ and if U is a co-ordinate neighborhood containing x and y then if we identify the co-ordinates on U with the co-ordinates they induce on T_x we have:

$$r^{m-h}(x, y)\, G(x, y) - r^{-h}(x, y)\, F_x(x - y) \in \Gamma(A_1^p \otimes A_2^p, U \times U)$$

where by $x - y$ we mean subtraction with respect to the co-ordinates on U.

We are concerned with the reduction of the singular integral equation $J\varphi = \alpha$ to a Fredholm equation. Let $\mathcal{S}^p$ be the set of all singular operators on differential forms of degree p; then we have

Definition 2: If $J \in \mathcal{S}^p$, the equation $J\varphi = \alpha$ is *regularizable* if there exists an operator $K \in \mathcal{S}^p$ such that the equation $K \circ J\varphi = K\alpha$ is a Fredholm equation which has the same solutions as the original equation.

By applying[2] the composition theorems of Giraud,[3] we obtain

THEOREM 1. *$\mathcal{S}^p$ is an algebra under operator composition. The subset $\mathfrak{F}^p$ of $\mathcal{S}^p$, consisting of those $J \in \mathcal{S}^p$ for which $\xi = 0$ and $F = 0$, is a two-sided ideal of $\mathcal{S}^p$.*

Now for $x \in M$ and $J \in \mathcal{S}^p$ we construct a singular operator $\boldsymbol{H}_x J$ on forms of degree p over T_x. The important property of $\boldsymbol{H}_x J$ is that J is regularizable if and only if $\boldsymbol{H}_x J$ has an inverse for each $x \in M$. $\boldsymbol{H}_x J$ is defined by

$$\boldsymbol{H}_x J\sigma(u) = \xi(x)\cdot\sigma(u) + \int_{T_x} \frac{\tilde{F}_x(u, v)}{\|u - v\|^m} \wedge * \sigma(v),$$

where $\tilde{F}_x(u, v) \in \Gamma(A_1{}^p \otimes A_2{}^p, T_x \times T_x - \{\text{diagonal}\})$ and is defined by

$$(\tilde{F}_x(u, v))(a_1, \ldots, a_p, b_1, \ldots, b_p) = (F_x(u - v))(a_1 - u, \ldots, a_p - u, b_1 - v, \ldots, b_p - v)$$

where $a_i, b_i \in T_x$. We denote by $\mathfrak{IC}_x{}^p$ the set of all operators on T_x which can be obtained from the mappings F_x as above. Now let $\mathfrak{F}_x{}^p$ be the ideal of $\mathcal{S}^p$ consisting of all operators with $\xi(x) = 0$ and $F_x = 0$; then we have

THEOREM 2. *$\mathfrak{IC}_x$ is an algebra under operator composition, and the map $\boldsymbol{H}_x: \mathcal{S}^p \to \mathfrak{IC}_x{}^p$ is a homomorphism of algebras with kernel $\mathfrak{F}_x{}^p$.*

By using the partition of unity, we obtain[2] the following theorem which shows that $\boldsymbol{H}_x$ is onto.

THEOREM 3. *If F is a bundle map satisfying conditions b and c and if $\xi \in \Gamma(A^p \oplus A^p, M)$, then there exists $J \in \mathcal{S}^p$ which corresponds to F and ξ.*

Since the space T_x has an inner product, we have an isomorphism $T_x \to T_x^*$ which induces a natural isomorphism between $A_x^p \otimes A_x^p$ and the space of linear transformations $A_x^p \to A_x^p$. This isomorphism provides $A_x^p \otimes A_x^p$ with an algebraic structure. Thus $\mathcal{S}_x^p$, the set of C^∞ maps $S_x \to A_x^p \otimes A_x^p$, is an algebra. Note that $\mathcal{S}_x^p$ can be represented as the algebra of matrices of C^∞ functions on S_x.

Consider the map $\boldsymbol{S}_x : \mathfrak{C}_x^p \to \mathcal{S}_x^p$ given by

$$(\boldsymbol{S}_x \boldsymbol{H}_x J)(u) = \xi(x) + \int_{T_x} e^{iu \cdot v} \frac{F_x(v)}{|v|^m}.$$

Then a generalization of the classical properties of Fourier transforms gives

THEOREM 4. *The map $\boldsymbol{S}_x : \mathfrak{C}_x^p \to \mathcal{S}_x^p$ is an isomorphism of algebras.*

We now express $\boldsymbol{S}_x$ in terms of spherical harmonics. If F_x is given by the matrix $(F^{ij}(u))$, $u \in S_x$, and if $\xi(x)$ is given by (ξ^{ij}), then we obtain

THEOREM 5. *If $F^{ij} = \sum_{n=1}^{\infty} Y_n{}^{ij}$ where $Y_n{}^{ij}$ is a spherical harmonic of degree n, then*

$$(\boldsymbol{S}_x \boldsymbol{H}_x J)(u) = \xi^{ij} + \sum_{n=1}^{\infty} \gamma_n Y_n{}^{ij}(u),$$

where

$$\gamma_n = i^n \pi^{m/2} \frac{\Gamma(n/2)}{\Gamma[(m+n)/2]}.$$

The analogue of the above theorem for singular integral equations on functions on the plane was first obtained by F. Tricomi.[4] S. G. Michlin[5] has improved Tricomi's proof and has made extensive use of the generalization of this result to n dimensions. This generalization was first stated by Giraud.[6] (See also Bochner,[8] Calderon, and Zygmund.[9])

To establish the above result, we apply[2] the Funk-Hecke[7] theorem to

$$(\boldsymbol{S}_x \boldsymbol{H}_x J)(u) = \xi^{ij} + \int_{S_x} R(u \cdot v) F_x(v),$$

where

$$R(u \cdot v) = \begin{cases} -\log |u \cdot v| + i\frac{\pi}{2} & \text{for } u \cdot v > 0, \\ -\log |u \cdot v| - i\frac{\pi}{2} & \text{for } u \cdot v < 0, \end{cases}$$

and we obtain

$$\gamma_n = \frac{w_{m-2}}{C_m^{1/2 m-1}(1)} \int_{-1}^{1} R(x)\, C_n^{1/2 m-1}(x)(1-x^2)^{1/2 m-\frac{3}{2}}\, dx,$$

where C_n^{ν} is the Gegenbauer polynomial[2] defined by

$$(1 - 2xt - t^2)^{\nu} = \sum_{n=0}^{\infty} C^{\nu}{}_n(x) t^n$$

and w_{n-2} is the volume of the unit sphere in Euclidean $(m-1)$-space.

Now we define the homomorphism $\boldsymbol{S}_M : \mathcal{G}^p \to \mathcal{S}_M{}^p$, where $\mathcal{S}_M{}^p$ is the algebra of bundle maps of $S \to A^p \otimes A^p$. $\boldsymbol{S}_M$ is defined by

$$\boldsymbol{S}_M J(u) = (\boldsymbol{S}_x \boldsymbol{H}_x J)(u), \qquad \text{where } u \in S_x \subset S.$$

Then

THEOREM 6. *The equation $J\varphi = \alpha$ can be regularized if and only if $\boldsymbol{S}_M J$ has an inverse in $\mathcal{S}_M{}^p$.*

In the above-mentioned paper by D. C. Spencer and the author the following operators are introduced:

$$d_\tau \varphi = {}^1/_2(\tau - \sqrt{-1})\partial\varphi + {}^1/_2(\tau + \sqrt{-1})\bar{\partial}\varphi$$

where φ is a differential form on a finite almost-complex Kähler manifold V and τ is a point on the Riemann sphere. We set $d_\infty = d$ and we have $d_{\sqrt{-1}} = \partial$. Now the Poincairé-Fredholm method for the solution of the Neumann problem for d_τ with the boundary conditions $n\varphi$ and $nd_\tau\varphi$ prescribed on bV, the boundary of V, leads to the introduction for $x \in V$ of the integral

$$A(x) = \int_{bV} [\mathfrak{d}_y g_0(x,y) \wedge {}^*\alpha(y) + g_0(x,y) \wedge {}^*\beta(y)]$$

where g_0 is a fundamental singularity (see Spencer[10]) for the operator $\square$, in some domain containing the closure of V. We evaluate the limits of $nA(x)$ and $nd_\tau A(x)$ as x approaches the boundary. Thus we arrive at a system of integral equations

$$J_\tau \varphi = \alpha$$

By calculating the high-order terms in the kernels of J_τ we find that J_τ is a singular operator for $\tau \neq \infty$ and an ordinary integral operator for $\tau = \infty$. Applying theorems 5 and 6 we obtain:

THEOREM 7. *The operator $J\tau$ is regularizable if and only if $\tau \in Z$, where* $Z = \{\tau \mid Re(\tau) = 0 \text{ and } Im(\tau) \leq 1\}$.

An example of K. Kodaira shows that for $\tau \in Z$ an eigen value of $J\tau$ can have infinite multiplicity.

* This note pertains to work done while the author was a National Science Foundation fellow.

[1] N. E. Steenrod, *Topology of Fibre Bundles* (Princeton, N.J.: Princeton University Press, 1951).

[2] J. J. Kohn, "A Non-self-adjoint Boundary Value Problem on Pseudo-Kähler Manifolds" (thesis, Princeton University, 1956).

[3] G. Giraud, "Équations à intégrales principales," *Ann. École norm. supér.*, ser. III, **51**, 231–372, 1934.

[4] F. Tricomi, "Equazioni integrali contenti il valor principale di un integrale doppio," *Math. Z.*, **27**, 87–133, 1928.

[5] S. G. Michlin, *Singular Integral Equations* (Am. Math. Soc. Translation No. 24 [New York, 1950]).

[6] G. Giraud, "Sur une classe générale d'équations à intégrales principales," *Compt. rend. Acad. sci.* (Paris), **202**, 2124–2127, 1936.

[7] H. Bateman, *Higher Transcendental Functions* (New York: McGraw-Hill Book Co., 1953).

[8] S. Bochner, *Theta Relations with Spherical Harmonics*, these PROCEEDINGS, **37**, 804–808, 1951.

[9] A. P. Calderon and A. Zygmund, "On the Existence of Certain Singular Integrals," *Acta Mathematica*, **88**, 85–139, 1952, and "On Singular Integrals," *Am. J. Math.*, **78**, 289–320, 1956.

[10] D. C. Spencer, "Potential Theory and Almost-Complex Manifolds," *Lectures on Functions of a Complex Variable*, Ann Arbor: Univ. of Michigan Press, pp. 15–43, 1955.

MEASURES OF THE VALUE OF INFORMATION

By John McCarthy

DARTMOUTH COLLEGE, HANOVER, NEW HAMPSHIRE

Communicated by Claude Shannon, July 12, 1956

1. *Introduction.*—Our knowledge of a future event may take the form of a set of probabilities $p_1, \ldots, p_n$. For example, we might have probabilities of $^3/_8$, $^1/_8$, and $^1/_2$ for rain, snow, and clear as tomorrow's weather. In communication theory our interest is in the various events only as carriers of a coded message. For this purpose Shannon's[1] entropy $-\sum p_i \log p_i$ is the appropriate measure of our uncertainty, and a function $A \sum p_i \log p_i + B$ is a good measure of what it is worth to be given these probabilities. In our weather example we care which event occurs. Furthermore, we may be more interested in whether the sky is clear than in whether rain or snow occurs if the weather is bad. In this paper we show that any convex function of a set of probabilities may serve as a measure of the value of information and that two such functions are equivalent in an appropriate sense if and only if they differ by a linear function.

2. *The Forecaster and His Client.*—We get our quantitative measures of the value of information from a situation in which a client pays a forecaster for predictions of a future event according to the following rules:

(i) The forecaster gives the client probabilities $q_1, \ldots, q_n$ for the events, where $\sum q_i = 1$.

(ii) The client takes action on the basis of these probabilities, and one of the possible events occurs.

(iii) If the ith event occurs, the client pays the forecaster $f_i(q_1, \ldots, q_n)$, which is abbreviated $f_i(q)$.

(iv) We assume that neither the forecaster nor the client can influence the predicted event, although the forecaster can make experiments to help predict it, and the client gets an amount which depends on both the action he takes and on the event which occurs. In what follows, it is assumed that the forecaster and the client both wish to maximize the expected value of their incomes.

Assuming that to the forecaster the probabilities of the possible events are $p_1, \ldots, p_n$, his expectation is $\sum p_i f_i(q)$ if he tells the client the q's. A payoff rule is said to "keep the forecaster honest" if, regardless of the value of $p = (p_1, \ldots, p_n)$, the forecaster's expectation is maximized if and only if he puts $q = p$, i.e., $q_i = p_i$ for each i.

Theorem 1. *A payoff rule keeps the forecaster honest if and only if* $f_i(q) = (\partial/\partial q_i) f(q)$, *where* $f(q)$ *is a convex function of* q *which is homogeneous of the first degree. The expectation of an honest forecaster is then* $\sum p_i f_i(p) = f(p)$.

We omit the proof. The derivative has to be taken in a suitable generalized sense. $f(q)$ is called a "payoff function" if it satisfies the conditions of Theorem 1.

I. J. Good[2] considered the problem of paying the forecaster with the restriction that $f_i(q) = F(q_i)$, i.e., the payoff depends only on the probability assigned to the event which actually occurred. He showed that putting $F(x) = A \log x + B$ keeps the forecaster honest, and Gleason (unpublished) showed that this is the only $F(x)$ which does. The forecaster's expectation is then $A \sum p_i \log p_i + B$, i.e., he is paid a fixed fee *minus* the expected uncertainty about the event after his prediction.

3. *The Client's Expectation.*—Suppose that on the basis of the forecaster's prediction the client chooses the jth of the actions open to him and that his payoff if the ith event occurs is a_{ij}. His expectation will be $g(p) = \max_j \sum_i a_{ij}p_i$ if j is chosen optimally.

From the theory of convex functions we have

THEOREM 2. *Any function $g(p)$ defined for $p_1 \geq 0, \ldots, p_n \geq 0$ which is convex and homogeneous of the first degree can be written in the form $\max_j \sum a_{ij}p_i$. Unless $g(p)$ is piecewise linear, there will have to be an infinite number of actions j.*

If we put $f(p) = g(p)$, the client is eliminated from the picture, since under this condition he turns all his gains over to the forecaster and is reimbursed for all his losses. This is not a satisfactory solution to the problem, so let us see what payoffs f are equivalent in their effect on the forecaster's efforts to get information.

4. *The Forecaster's Experiments.*—Assume that the forecaster has a priori probabilities $r_1, \ldots, r_n$ for the events, that he has a choice of m experimental procedures with expected costs to him of $c_1, \ldots, c_m$, and that the conditional probability of the kth outcome of the hth experiment given that the ith event will occur is s_{khi}. The experiment chosen by the forecaster will depend on the c's, the s's, and the r's and on the payoff function chosen by the client. We call two payoff rules equivalent if, for any set of c's, s's, and r's, they lead to the same choice of experiment by the forecaster.

THEOREM 3. *$f(q)$ and $f^*(q)$ are equivalent if and only if $f(q) = f^*(q) + \sum a_i q_i$, i.e., if the two payoff functions differ by a linear function of the q's.*

The proof is omitted. If f and f^* are equivalent, then $f_i(q) = f_i^*(q) + a_i$, so that the payoff rules differ by an amount which depends only on the event which occurs and not on the forecaster's prediction. The forecaster's and client's interests will be identical if we put $f(q) = g(q) + \sum a_i q_i$. The a_i's are subject to negotiation between the client and the forecaster, and they determine both a base level of payment and also a betting relation between the client and forecaster. If f is normalized so that $f(1, 0, \ldots, 0) = f(0, 1, \ldots, 0) = \ldots$, the payment for a precise correct prediction is independent of the event predicted.

5. *Conclusion.*—The foregoing analysis shows that any convex function of a set of probabilities will, under appropriate circumstances, be a measure of the value of the information contained in a set of probabilities in the sense that it is an appropriate payment to a forecaster who furnishes the probabilities.

The intuitive content of the convexity restriction is that it is always a good idea to look at the outcome of an experiment if it is free. For suppose that the experiment has two outcomes, A and A^*, which would give one probabilities p and p^* for the event in question. Let t be the probability that A is the outcome. If we decide not to look, our expectation is $f(tp + (1 - t)p^*)$, while if we decide to look, our expectation is $tf(p) + (1 - t)f(p^*)$.

Finally, we remark that there are yet more general ways of paying the forecaster. For example, the client may agree to pay a certain fraction α of the costs of experimentation. Then the payoff function can be scaled down by a factor α with the identity of interests still preserved. We hope to treat these matters on another occasion.

[1] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Urbana: University of Illinois Press, 1949).

[2] I. J. Good, "Rational Decisions," *J. Roy. Stat. Soc., B*, Vol. **14**, No. 1, 1952.

DIOPHANTINE EQUATIONS IN CERTAIN RINGS

By H. S. Vandiver*

UNIVERSITY OF TEXAS

Communicated July 12, 1956

Let p be an odd prime and $F[p^n]$ denote a finite field of order p^n, and consider the solutions of the conditional polynomial equation

$$f(x_1, x_2, \ldots, x_s) = 0, \tag{1}$$

where the coefficients are in $F[p^n]$ and $s > 1$ if the equation contains a non-zero constant term; $s > 2$, if such is not the case. An equation which has been extensively studied, particularly in recent years, is the special case

$$c_1 x_1^{a_1} + c_2 x_2^{a_2} + \ldots + c_s x_s^{a_s} + c_{s+1} = 0, \tag{2}$$

where the a's are integers such that $0 < a < p^n - 1$; $s > 1$ for $c_{s+1} \neq 0$ and $s > 2$ for $c_{s+1} = 0$, the c's being given elements of $F[p^n]$ and

$$c_1 \ldots c_s \neq 0. \tag{2a}$$

In the present paper we consider the application of the theory of the equation (2) to the derivation of conditional equations with coefficients in R which have no solutions in R. Here R denotes a commutative ring with a unity element. When R is the rational ring, it was shown in several other papers[1, 2] how to derive certain classes of equations of this type, and in the present paper we shall use extensions of these methods.

In the case of the applications to the rational ring, we first employed the congruence

$$c_1 x_1^{a_1} + c_2 x_2^{a_2} + \ldots + c_s x_s^{a_s} + c_{s+1} \equiv 0 \pmod{p}, \tag{3}$$

with the c's rational integers, $c_1 \ldots c_s \not\equiv 0 \pmod{p}$; $s > 1$ for $c_{s+1} \not\equiv 0 \pmod{p}$ and $s > 2$ for $c_{s+1} \not\equiv 0 \pmod{p}$. Now we let R be a commutative ring containing a unity element and, analogous to equation (2), consider the congruence, in the unknowns $\omega_1, \omega_2, \ldots, \omega_s$, each ϵR,

$$\sigma_1 \omega_1^{a_1} + \sigma_2 \omega_2^{a_2} + \ldots + \sigma_s \omega_s^{a_s} + \sigma_{s+1} \not\equiv 0 \pmod{\mathfrak{p}}, \tag{4}$$

where $\mathfrak{p}$ is a prime ideal in R of definite norm p^n, p an odd prime, the norm of $\mathfrak{p}$ is the number of distinct elements in the residue class ring of $\mathfrak{p}$, and

$$\sigma_1 \sigma_2 \ldots \sigma_s \omega_1 \omega_2 \ldots \omega_s \not\equiv 0 \pmod{\mathfrak{p}}. \tag{4a}$$

Also, $s > 2$ if $\sigma_{s+1} = 0$, and $s = 2$ if $\sigma_{s+1} \neq 0$. The ideal $\mathfrak{p}$ being a prime ideal in R, the residue classes modulo $\mathfrak{p}$ form a finite field $F[p^n]$, since a finite integral domain is a field. Suppose that it is possible to show that congruence (4) has no solutions which satisfy condition (4a). Assume, further, that it has been determined that (4) has no solutions even when we allow any of the x's to be $\equiv 0 \pmod{\mathfrak{p}}$, aside from the trivial case where every x has this property, with $\sigma_{s+1} \equiv 0 \pmod{\mathfrak{p}}$. Now consider the Diophantine equation in R,

$$\gamma\sigma_{s+1} + \beta_{s+1} + \sum_{i=1}^{s} (\beta_i + \gamma\sigma_i)\rho_i^{a_i} = 0, \tag{5}$$

where $\beta_i \equiv 0 \pmod{\mathfrak{p}}$, $i = 1, 2, \ldots, s+1$, $\gamma \not\equiv 0 \pmod{\mathfrak{p}}$. Modulo $\mathfrak{p}$, this reduces first to

$$\gamma\sigma_{s+1} + \sum_{i=1}^{s=1} \gamma\sigma_i\rho_i^{a_i} \equiv 0 \pmod{\mathfrak{p}}, \tag{5a}$$

and since $\gamma \not\equiv 0 \pmod{\mathfrak{p}}$, we obtain relation (4), which we are assuming has no solutions; hence equation (5) has no solutions in the γ's ϵR.

In what follows, we shall show that this principle may be applied to many classes of Diophantine equations of type (5) and also related types.

If the laws of addition and multiplication are defined in R, then, if we except the zero residue class modulo $\mathfrak{p}$, the other residue classes form a cyclic group of order $p^n - 1$ under multiplication generated by an element $\theta \epsilon R$, and θ may be determined in a finite number of steps.

An example of the type of ring we have called R in the above is the ring of integers in an algebraic field, since in this ring every ideal has a finite norm.

As the residue classes modulo $\mathfrak{p}$ in congruence (4) form, as we have noted, a finite field, we may replace congruence (4) by an equation in a finite field such as (2) and study the number of solutions of (2) to obtain the number of solutions of congruence (4). Assuming that the x's in equation (2) are nonzero in $F[p^n]$, in another paper[3] we employed g, a generator of the cyclic group formed by the nonzero elements in the finite field $F[p^n]$; if $c_{s+1} \neq 0$, we replaced equation (2) by

$$1 + E(e_1 + \alpha_1 a_1) + \ldots + E(e_s + \alpha_s a_s) = 0, \tag{6}$$

where $E(t) = g^t$, and if $c_{s+1} = 0$, equation (2) was replaced by

$$E(e_1 + \alpha_1 a_1) + \ldots + E(e_s + \alpha_s a_s) = 0, \tag{6a}$$

provided that in equation (2) $x_i \neq 0$, $i = 1, 2, \ldots, s$. Let $(p^n - 1, a_i) = d_i$, $i = 1, 2, \ldots, s$. Now the α_i in equations (6) and (6a) may be taken in the set

$$0, 1, 2, \ldots, \left(\frac{p^n - 1}{d_i}\right) - 1, \qquad i = 1, 2, \ldots, s.$$

The number of solutions of equation (2) for nonzero x's is equal to the number of sets of values $\alpha_1, \alpha_2, \ldots, \alpha_s$ in equations (6) and (6a). Formulas were employed which showed that this number of solutions could be determined[3] (p. 50) directly if the number of sets of the solutions α_1, α_2 in $F[p^n]$ of

$$E(u_1 + \alpha_1 m) + E(u_2 + \alpha_2 m) + 1 = 0 \tag{7}$$

was known for any integers u_1 and $u_2 \geqq 0$, where m is the L.C.M. of $d_1, d_2, \ldots d_s$.

As m is the L.C.M. of the a's in congruence (4), the simplest case to consider appears to be the case when the a's are all equal. In an earlier paper,[4] referred to above, this case was considered, and also the case when R is a rational ring and the congruences have no constant term. The results were extended by Ankeny.[5] The methods employed in these two papers do not, however, apply to Diophantine

equations with non-zero constant term, whereas the methods we are describing here do apply, to some extent, to this case.

It is possible to obtain non-solvable sets of equations of type (7) with little of the element of trial involved. If we consider the equation

$$\alpha_1 x_1^m + \beta_2 x_2^m + \gamma_3 = 0,$$

with α_1, β_2, and γ_3 each ϵR and the corresponding congruence

$$1 + \alpha_1 z_1^m \equiv \beta_1 z_1^m \pmod{\mathfrak{p}}, \tag{9}$$

then it is known that if we use the last congruence in the form, equivalent to (7),

$$1 + \theta^{a+rm} \equiv \theta^{b+sm} \pmod{\mathfrak{p}}, \tag{10}$$

with θ defined as before, and if $[a, b]_m$ denotes the number of solutions (r, s) in equation (8), then, if $p^n = 1 + dm$ and if $d < m$, we may use the known relation

$$\sum_{a=0}^{m-1} [a, b]_m = d \quad \text{or} \quad d - 1, \tag{11}$$

so that at least one of the $[a, b]$'s must be zero for each b.

To examine the above possibility, we may set up a table giving the values of the $[a, b]$'s in (10) for each a and b; then, to obtain the number of solutions of congruence (4) with each $\omega \not\equiv 0 \pmod{\mathfrak{p}}$, we employ the formulas[6] which give the number of solutions of (4) under conditions (4a). (This method had already been described in the second reference in our earlier paper[1] for the case of R, the ring of rational integers.) In the case in which the number of solutions turn out to be zero, we may have an application to Diophantine equations in R, if we find, upon examining further the number of solutions of (4), where some of the ω's may be $\equiv 0 \pmod{\mathfrak{p}}$, is zero. This requires the consideration of a set of congruences obtained from (4) by substituting zero for each ω in any combination of ω's and determining the number of solutions, with no ω congruent to 0 (mod $\mathfrak{p}$), of each of the resulting congruences in the various ω's remaining. If it turns out that none of these congruences have any solutions in R, except for the trivial case $\sigma_{s+1} \equiv 0 \pmod{\mathfrak{p}}$, and all the ω's are zero, then it follows that none of equations (5) have any solutions. *As we see from the proof of Theorem III of a previous paper,[4] when we were considering the equation in the ring of rational integers, and also relation* (11) *of our present paper, we are not likely to be able to set up many types of equations without solutions of type* (5), *with each* $a = m$, *unless in the relation* $p^n - 1 = dm$, m *is quite large with respect to* d.

Elsewhere [1, 2] the congruence

$$1 + ax^c \equiv by^l \pmod{p} \tag{11a}$$

was treated in the case in which it is conditional in x and y, p is an odd prime, $ab \not\equiv 0 \pmod{p}$, and l and c are integers each greater than 1 such that $lc = p - 1$. By setting up tables[1, 2] of the possible solutions of these congruences for a given p, it was shown how to employ these to derive nonsolvable equations of the type

$$1 + ax_1^c = by_1^l. \tag{11b}$$

In a similar way we may use the congruence

$$1 + \delta_1\mu_1^r = \delta_2\mu_2^l \pmod{\mathfrak{p}}, \tag{12}$$

where δ_1, δ_2, μ_1, μ_2 are elements in R, and $\mathfrak{p}$, as before, is a prime ideal in R such that $N(\mathfrak{p}) = p^n = 1 + cl$, with θ defined in R as before and assuming

$$\delta_1\delta_2\mu_1\mu_2 \not\equiv 0 \pmod{\mathfrak{p}}. \tag{13}$$

We may replace (12) by

$$1 + \theta^{i+cr} \equiv \theta^{j+ls} \pmod{\mathfrak{p}}, \tag{14}$$

with r in the set $0, 1, \ldots, l-1$; s in the set $0, 1, \ldots, c-1$, i and j fixed with $0 \leqq i \leqq c-1$; $0 \leqq j \leqq l-1$. The equivalence of the two congruences (12) and (14) follows easily from the properties of the θ's. We shall denote the number of solutions r, s in the last congruence within the limits mentioned $[i, j]_{cl}$. Since the residue classes in R modulo $\mathfrak{p}$ form a finite field, we may use the following known relations: If l is odd,

$$\sum_{j=0}^{l-1} [i, j]_{cl} = l - \begin{cases} 0 & \text{if } i \neq \dfrac{c}{2} \\ 1 & \text{if } i = \dfrac{c}{2} \end{cases} \tag{15}$$

and if l is even,

$$\sum_{j=0}^{l-1} [i, j]_{cl} = l - \begin{cases} 0 & \text{if } i \neq 0, \\ 1 & \text{if } i = 0, \end{cases} \tag{16}$$

and also

$$\sum_{i=0}^{c-1} [i, j]_c = l - \begin{cases} 0 & \text{if } j \neq 0, \\ 1 & \text{if } j = 0. \end{cases} \tag{17}$$

Also, in testing (14), suppose that we find, for a particular i, j, that $[i, j] = 0$. To apply this result in order to obtain Diophantine equations with no solutions, we have to test the possibility that (12) is solvable for either $\mu_1 \equiv 0 \pmod{\mathfrak{p}}$ or $\mu_2 \equiv 0 \pmod{\mathfrak{p}}$, assuming

$$\delta_1 \equiv \theta^i \quad \text{and} \quad \delta_2 \equiv \theta^j \pmod{\mathfrak{p}}$$

in (12). For this we have the following criterion:

If $(\mu_1, p) = p$, $(\mu_2, p) = 1$, then (12), where $p^n - 1 = cl$, $i = 0, 1, \ldots, c-1$, and $j = 0, 1, \ldots, l-1$, has solutions if and only if $j = 0$. If $(\mu_2, p) = p$, and $(\mu_1, p) = 1$, then (12), $i = 0, 1, \ldots, c-1$, and $j = 0, 1, \ldots, l-1$, has solutions when l is odd if and only if $i = c/2$; when l is even, if and only if $i = 0$.

Based on the tables of the $[i, j]$'s calculated,[1, 2] about 25 per cent of the congruences (11a) for a given p and l led to impossible equations of type (11b).

In order to apply certain formulas which have been published elsewhere, we employ (14) in the form

$$\theta^{i_1+cs_1} + \theta^{j_1+lt_1} \equiv 1 \pmod{\mathfrak{p}}. \tag{18}$$

If $(i_1, j_1)_{cl}$ denotes the number of solutions s_1, t_1 in the above, then, obviously,

$$(i_1 + \epsilon, j_1)_{cl} = [i, j]_{cl}, \tag{18a}$$

where $\epsilon = (p^n - 1)/2$.

We shall now examine the question as to the number of (i, j)'s which are zero for a given p, c, and l. To effect this, we shall employ the formula[4] if $b \not\equiv 0 \pmod{c}$, with $lc + 1 = p^n$, $v = p^n - 1$,

$$\sum_{1=0}^{l-1} (b, i)_{cl}^2 = \sum_{\substack{j=0 \\ d}}^{c-1} (d, j)_{vc}(d, j - b)_c + l, \tag{19}$$

where in the summation on the right-hand side of this equation d ranges over the multiples of l in the set $0, 1, \ldots, v - 1$. Let l be odd; then, in view of (15) and (18a) we see that for $i = 0$ we have

$$\sum_{i=0}^{l-1} (0, j)_{cl} = l - 1. \tag{19a}$$

Consequently, at least one of the summands on the left equals zero. For the cases where $b \not\equiv 0 \pmod{c}$, we may use formula (19), and if we assume that none of the (i, j)'s under this assumption is zero, it follows that each must equal unity; hence (19) gives

$$(d, h)_{vc}(d, h - b)_c = 0 \tag{20}$$

for each h in the set $0, 1, \ldots, c - 1$, and d is any multiple of l in the set $0, 1, \ldots, v - 1$; b any integer in the set $1, 2, \ldots, c - 1$.

In equation (20) for a given d there is just one $(d, h')_{vc}$ which equals unity; hence (20) gives

$$(d, h' - b)_c = 0 \tag{21}$$

for $b = 1, 2, \ldots, c - 1$. Then, if we assume $d \not\equiv 0 \pmod{c}$ and employ (17), we have

$$\begin{aligned} \sum_{b=1}^{c-1} (d, h_1 - b)_c &= l - (d, h_1')_c \\ &= l - 1, \end{aligned}$$

which contradicts (21). The above argument was made under the restriction that l was odd. If it is even, the same proof goes through with slight modification. Hence we conclude that there are always two of the $(i, j)_{cl}$'s that are equal to zero, and consequently also two of the $[i, j]$'s. However, induction indicates in the case of $n = 1$ that about one-third of the $[i, j]$'s are zero.[2]

Elsewhere[1, 2] we have mentioned the problem of determining the primes p in (14) with $p = lc + 1$ such that $[i, j] = l$ for some i and j with l fixed, which we called "Case A" of (14). Also, we noted the existence of seven primes p with $l = 5$ such that $[i, j] = 5$. But we noted that for $l = 7$ there was no prime $p < 2003$ such that Case A held for any particular i and j. *It may happen, of course, that since c increases with p, for p sufficiently large we shall always find an $[i, j] = l$.* On the other hand, we know that,[7] if in Case A, $[b, h]_{cl} = l$ for some h with a fixed $b \not\equiv 0 \pmod{c}$, l odd,

$$\left[d + \frac{c}{2}, i\right]_{vc} \left[d + \frac{c}{2}, i - b\right]_{c} = 0,$$

for $d \not\equiv 0 \pmod{l}$, with d in the range $0, 1, \ldots, v-1, v = p-1$, i any integer in the set $0, 1, \ldots, c-1$. *So the number of these criteria also increases with p.* For the case when $l = 3$, Case A occurs when $p = 103, 109, 127, 139, 163$, and it occurs more than once for each of these primes. However, the case $l = 3$ may be exceptional, at least as far as numerous occurrences of Case A are concerned. Often, in dealing with trinomials, if we are using some prime parameter in our investigation of them, the prime 3 is exceptional.

In other papers the writer has used the term "Case B" to indicate the case when, for i fixed, each one of the $(i, j)_{cl}$'s is unity. This situation has already arisen in the present paper in connection with (18), and in view of application of (19) when each $(b, i) = 1$ (and therefore for each $[b, i]$), we may then state:

If b is fixed and $\not\equiv 0$ *(mod c), then, if Case B holds, relations* (20) *follow.*

There is a connection between Cases A and B of the $[i, j]_{cl}$'s, which we shall now explain. Consider the cyclic group formed by the θ's in (18). If c is prime to l, it may be represented as the product of two cyclic groups of orders c and l, generated by θ^l and θ^c. In view of this, we may replace equation (14) by

$$1 + \theta^{al}\theta^{ct} \equiv \theta^{bc}\theta^{lu} \pmod{\mathfrak{p}}, \tag{22}$$

with $i_1 \equiv al \pmod{c}$, $j_1 \equiv bc \pmod{l}$, where t is in the range $0, 1, \ldots, l-1$, and where u is in the range $0, 1, \ldots, c-1$. Clearly $[i, j]_{cl} = [al, bc]_{cl}$. Suppose that $[al, bc]_{cl} = d$ in congruence (22); in view of this we are led to the congruences

$$1 + \theta^{al}\theta^{ct_k} \equiv \theta^{bc}\theta^{lu_k} \pmod{\mathfrak{p}}, \tag{23}$$

where $k = 1, 2, \ldots, d$, and no two of the t's are congruent, modulo l. Division gives

$$1 + \theta^{-al}\theta^{-ct_k} \equiv \theta^{(b-t_k)c}\theta^{l(u_k-a)} \pmod{\mathfrak{p}}. \tag{23a}$$

Now in the exponent $(b - t_k)c$ on the right, since the t's are incongruent modulo l, then so are the $(b - t_k)$'s. We see from this that the value $[-al, bc]_{cl}$ may be, in general, quite different from $[al, bc]$. In particular if $d = k = l$, then the $(b - t_k)$'s are congruent in some order, modulo l, to $0, 1, \ldots, l-1$, so that we have:

If Case A holds for a given i in equation (14), *then Case B holds for* $(-i)$ *in lieu of i, if c is prime to l.*

The converse of this statement does not hold. The argument just given, however, also shows that Case A never arises for $i = 0$ in equation (14).

For $c \equiv 0 \pmod{l}$, we note that (14) gives

$$1 + \theta^{-i}\theta^{-cr} \equiv \theta^{j-t}\theta^{ls_1} \pmod{\mathfrak{p}},$$

and it is easy to see then that the result stated above holds in the case $c \equiv 0 \pmod{l}$ also. The situation when $c \equiv 0 \pmod{l_1}$, $1 < l_1 < c$, is more complicated.

We have already referred in this paper to the application of a nonsolvable congruence of type (11*a*) to deriving nonsolvable equations of type (11*b*). In the case in which $l = 2$, these ideas may be put in a very simple form. We shall consider the congruence

$$a + bx^{\mu} \equiv ky^2 \pmod{p}; \qquad \mu = \frac{p-1}{2}. \tag{24}$$

If $xy \not\equiv 0 \pmod{p}$, $abk \not\equiv 0 \pmod{p}$, and, using Legendre symbols,

$$\left(\frac{(a-b)k}{p}\right) \neq 1, \qquad \left(\frac{(a+b)k}{p}\right) \neq 1, \tag{25}$$

then (24) has no solutions, under the conditions mentioned, as we see by using $x^{\mu} \equiv \pm 1 \pmod{p}$. If we now consider the possibility $xy \equiv 0 \pmod{p}$, this is excluded if we assume

$$\left(\frac{ak}{p}\right) \neq 1, \tag{26}$$

and

$$a \pm b \not\equiv 0 \pmod{p}. \tag{27}$$

Consequently, under conditions (25), (26), and (27), congruence (24) has no solutions whatever in integers. Hence the equation

$$a + bx_1^{\mu} = ky_1^2 \tag{28}$$

has no solutions in integers.[8]

Assuming p, l, and c fixed in congruence (24), we may take a, b, and c each in the set $0, 1, \ldots, p-1$, and classify all the impossible congruences derivable by this method, in a finite number of steps. This is another form of the process we used of setting up tables for a given p of the impossible congruences of the type (24), except that, in the second process, primitive roots of p were employed; but this is not necessary in the first method, in which we may employ the law of quadratic reciprocity conveniently.

We will now consider congruence (24) again, but from a different point of view. We can regard a, b, and k as fixed and let p vary. We can then note that there is an infinity of primes p fitting conditions (25) and (26), with a, b, and k fixed and provided that none of the quantities

$$(a-b)k, \qquad (a+b)k, \qquad ak \tag{29}$$

is the square of an integer and also that the product of no two of these quantities has this property, as well as the product of all three.[9] As far as I know, the known proofs of this result are not constructive, and we do not know of any way of determining such primes in general; but for some small given values of a, b, and k we may determine some primes of the type desired, using the law of quadratic reciprocity.

The ideas just explained may be extended to the examination of the congruence with $(a, b) = 1$, a and b integers,

$$a + bx^c \equiv y^l \pmod{p}, \tag{30}$$

with p an odd prime, likewise l, and the equation

$$a + bx^c = y^l, \tag{31}$$

a, b, and l given, $ab \not\equiv 0 \pmod{\mathfrak{p}}$.

Set, if $N(\mathfrak{p})$ is the norm of $\mathfrak{p}$,

$$\left(\frac{\omega}{\mathfrak{p}}\right)_l = \zeta^{(N(\mathfrak{p})-1)/l} \pmod{\mathfrak{p}},$$

where ω is an integer $\epsilon K(\zeta)$, an algebraic field defined by $\zeta = e^{2i\pi/l}$, $\mathfrak{p}$ is a prime ideal in K prime to (ωl) and which divides (p). By a theorem given by Takagi,[10] there exists an infinity of prime ideals $\mathfrak{p}$ of the first degree ϵK such that, if $(a^l + b^l)$, and $(\mathfrak{p}l)$ are prime to each other,

$$a + b\zeta^i = \alpha_i, \qquad i = 0, 1, \ldots, l-1,$$

then

$$\left(\frac{\alpha_0}{\mathfrak{p}}\right)_l = \zeta^{a_0 m}, \left(\frac{\alpha_1}{\mathfrak{p}}\right)_l = \zeta^{a_1 m}, \ldots, \left(\frac{\alpha_{l-1}}{\mathfrak{p}}\right)_l = \zeta^{a_{l-1} m}, \tag{32}$$

where the a's are prescribed integers in the set $0, 1, \ldots, l-1$, and m is some integer $\not\equiv 0 \pmod{l}$, provided that, if the relation

$$\alpha_0^{m_0}\alpha_1^{m_1} \ldots \alpha_{l-1}^{m_{l-1}} = \gamma^l, \tag{33}$$

with γ an integer in K, and the m's integers, then $m_i \equiv 0 \pmod{l}$, $i = 0, 1, \ldots,$ holds, $l-1$. We then select $\mathfrak{p}$ so that $ab \not\equiv 0 \pmod{\mathfrak{p}}$.

We now consider congruence (30) and assume in equation (32) that none of the a_i's, $i = 0, 1, \ldots l-1$, is divisible by l. Let θ be the generator of the cyclic group of the residue classes (excluding the zero class) modulo $\mathfrak{p}$ under multiplication. Then we may select θ so that $\theta^c \equiv \zeta \pmod{\mathfrak{p}}$. Hence we may obtain from congruence (30)

$$a + b\theta^{hc} \equiv y^l \pmod{\mathfrak{p}}, \tag{34}$$

$$a + b\zeta^h \equiv y^l \pmod{\mathfrak{p}}, \tag{34a}$$

whence

$$\left(\frac{a + b\zeta^h}{\mathfrak{p}}\right)_l = 1,$$

for some h, if congruence (34) is satisfied. However, if we select, as above, the α's to satisfy equations (32) with none of the a's $\equiv 0 \pmod{l}$, this shows that congruence (30) is impossible for $xy \not\equiv 0 \pmod{l}$. As to the last restriction, this can be removed if we assume that

$$\left(\frac{a}{\mathfrak{p}}\right)_l \neq 1; \qquad a^l + b^l \not\equiv 0 \pmod{\mathfrak{p}}, \tag{35}$$

which is always possible if a is not the lth power of an integer; consequently equation (31) has no solutions, provided that we can show that $m_i \equiv 0 \pmod{l}$, $i = 0, 1, \ldots, l-1$, in equation (33) and also that conditions (35) are satisfied. As to the first condition, we note that $a + b\zeta^i$ is prime to $a + b\zeta^j$, $i \not\equiv j \pmod{l}$, as, if they have a common prime ideal factor $\mathfrak{q}$, then, if $\mathfrak{a}_i$ is an ideal in K

$$b(\zeta^i - \zeta^j) \equiv 0 \pmod{\mathfrak{q}},$$

or $\mathfrak{q}$ divides (l), contradicting $(a^l + b^l, l) = 1$, noting also that $b \not\equiv 0 \pmod{\mathfrak{p}}$ is impossible, since $(a, b) = 1$. Hence equation (33) gives

$$(a + b\zeta^i)^{m_i} = \mathfrak{a}_i^l, \qquad i = 0, 1, \ldots, l - 1,$$

by the unique factorization of ideals.

If $m_i \not\equiv 0 \pmod{l}$, set $m_i m_i' = 1 + kl$; then

$$\begin{aligned}(a + b\zeta^i)^{m_i m_i'} &= (a + b\zeta^i)^{1+kl}\,\mathfrak{a}_i \\ &= (a + b\zeta^i)\,\mathfrak{a}_i^l,\end{aligned}$$

or

$$(a + b\zeta^i) = \mathfrak{b}_i^l. \tag{36}$$

If $i \not\equiv 0 \pmod{l}$ in equation (36), we may replace ζ by ζ^s for $s = 1, \ldots, l - 1$, and multiply together the resulting equations, to obtain, if $a \neq v^l$,

$$\frac{a^l + b^l}{a + b} = d^l. \tag{37}$$

Hence $m_1, m_2, \ldots m_{l-1}$ are all zero unless equation (37) holds. If $m_0 \neq 0$, then we have

$$a + b = t^l. \tag{38}$$

It then follows that unless one of the two equations (37) or (38) is satisfied and every m in (33) is zero, we may select the a's in equation (32) to be all prime to l. We then have the following theorem.

THEOREM. *The equation*

$$da + k_1 p + (db + k_2 p)x^c = (d + k_3 p)y^l, \tag{39}$$

with $(a, b) = 1$, a, b, *and* l *fixed, is impossible for an infinity of integers* c *such that* $cl + 1 = p$, *with* p *an odd prime provided that* l *is an odd prime,*

$$\frac{a^l + b^l}{a + b} \neq r^l; \qquad a + b \neq s^l; \qquad a \neq t^l, \tag{40}$$

and a, b, r, s, t *are integers,* $d \not\equiv 0$ *(mod* p*).*

* Part of the work on the present paper was done under National Science Foundation Grant G1397.

[1] Erna H. Pearson and H. S. Vandiver, "On a New Problem concerning Trinomial Congruences Involving Rational Integers," these PROCEEDINGS, **39**, 1278–1285, 1953.

[2] J. L. Selfridge, C. A. Nicol, and H. S. Vandiver, "On Diophantine Equations Which Have No Solutions," these PROCEEDINGS, **42**, 264–266, 1956.

[3] H. S. Vandiver, "On the Number of Solutions of Some General Types of Equations in a Finite Field," these PROCEEDINGS, **32**, 47–52, 1946.

[4] H. S. Vandiver, "On Classes of Diophantine Equations of Higher Degrees Which Have No Solutions," *ibid.*, pp. 101–106, 1946.

[5] N. C. Ankeny, "The Insolubility of Sets of Diophantine Equations in the Rational Numbers," these PROCEEDINGS, **38**, 880–884, 1952.

[6] Cf. these PROCEEDINGS, **32**, 50, 1946, relation (11), for $\sigma_{s+1} \neq 0 \pmod{\mathfrak{p}}$, and *ibid.*, p. 51, the paragraph immediately preceding the statement of Theorem I, which gives an explanation of the procedure in obtaining a formula providing for the case $\sigma_{s+1} \equiv 0 \pmod{\mathfrak{p}}$. The number of solu-

tions of (10) for any a and b, which is written here in a form more suitable for setting up tables, may easily be shown to give directly the number of solutions of (7).

[7] H. S. Vandiver, "On Trinomial Equations in a Finite Field," these PROCEEDINGS, **40**, 1008–1010, 1954, relation (12).

[8] We think it probable that the method discussed here is known, particularly in the case $l = 3$, but we have been unable as yet to locate it in the literature.

[9] Hilbert, *Gesammelte Abhandlungen*, **1**, 182–183 (Berlin: Julius Springer, 1932).

[10] T. Takagi, "Ueber eine Theorie des relativ-Abelschen Zahlkörpers," *J. Coll. Sci., Imp. Univ. Tokyo*, **41**, No. 9, 16, 1920.

A FUNDAMENTAL THEORY OF SUPERCONDUCTIVITY

BY KENNETH S. PITZER

DEPARTMENT OF CHEMISTRY AND CHEMICAL ENGINEERING, UNIVERSITY OF CALIFORNIA, BERKELEY, CALIFORNIA

Communicated June 6, 1956

Although phenomenological theories of superconductivity are now well developed, there is still need for an adequate fundamental theory based directly on the quantum behavior of the electrons and the lattice. Recently Schafroth[1] has shown that a Bose gas of charged particles would show the Meissner effect. Also, Bardeen[2] has shown that a separate particle system where all excited electronic levels are separated from the "surface of the Fermi sea" by a finite energy will show the Meissner effect.

Our analysis of the problem will be more descriptive than mathematical and will be similar in many respects to the theory of liquid helium II.[3]

The key feature is the finite binding energy of electron pairs or, more precisely, the finite energy required to excite the electronic system in modes other than translational flow of the entire system.

I

Our general knowledge of chemical substances leads us to expect electron pairing. In molecules chemical bonding is strongest for completely paired electrons in all but the rarest cases (such as O_2). We normally find unpaired electrons associated only with atomic inner shells, which are not involved in bonding such as in the rare earths. Metals at room temperature are indeed an exception to this general situation (although the fraction of unpaired electrons is infinitesimal). Simple theories of metals which are successful in many other respects fail to predict superconductivity; however, these theories fail to take adequate account of coulombic repulsion of electrons and make the assumption of a perfect lattice of stationary nuclei.

It is known[4] that lattice distortions tend to produce breaks in the energy-level bands. If a single distortion lowers the electronic energy sufficiently (more than the energy of zero-point vibration in the corresponding mode), then permanent distortion occurs. Even for a lattice which is perfect in the mean location of the atoms, however, the total energy must be averaged over the various distortions which arise from zero-point vibration. Thus, while the electronic state with com-

plete pairing of spins may be stable by only an infinitesimal margin for a perfect lattice, this margin may be expected to become finite when averaged over the distortions of the lattice.

Let us also discuss this question in terms of Pauling's[5] theory of metallic binding. One notes that there are a number of locations in the lattice that are appropriate for chemical bonds, i.e., the shortest distances occurring between atoms. In a metal the number of these bond sites exceeds the number of bonding electron pairs. This is in contrast to a crystal such as diamond where the number of sites is precisely equal to the number of bonding electron pairs.

In addition to the bond sites, one must consider the available atomic orbitals suitable for bonding electrons. While almost always a pair of suitable orbitals can be found for each bond site, there are limitations on the number and arrangement of orbitals which may be occupied simultaneously. Thus in graphite, while the 2p orbital perpendicular to the plane is appropriate for the formation of a π bond in any one of the three bond sites adjacent to the atom, only one of these bonds can be formed at a given time. This limitation is important and, while less restrictive in true metals, must be kept in mind. It will be more restrictive in metals where the ratio of electrons to valence-shell orbitals is high. Many such metals show superconductivity, while those with a low ratio of electrons to orbitals do not.

Pauling now constructs wave functions in which the electrons occupy these bond sites and bond orbitals either in pairs or singly. Coulombic repulsion between electrons is very strong; consequently the set of sites occupied in a low-energy wave function must maintain uniform charge density. The true wave function is taken as the optimum linear combination of these component functions (in the manner of quantum-mechanical resonance). The lowest energy is obtained if the electrons, which occupy orbitals singly in one component function, have spins paired, because the resonance may then include other functions with these electrons in a single bond orbita[1].

Such a system of electrons is free to move collectively through the lattice and thus conduct electricity. Symbolically one may write a possible wave function for a flowing metallic electronic system:

$$\Psi = \Phi \exp i\boldsymbol{k}\cdot(\sum_i \boldsymbol{R}_i), \tag{1}$$

where Φ is the wave function of the system in the absence of flow and the $\boldsymbol{R}_i$ vectors locate the electrons. Feynman[3] discusses this type of function more fully in connection with He II. The question of resistance, i.e., energy loss, from such a flow will be considered later.

Now let us recognize that the lattice is vibrating with zero-point amplitude supplemented by any lattice thermal excitation. As the lattice oscillates, some bond sites will become of more nearly ideal length than others. Also, the angles change, and it will be more favorable in terms of atomic orbitals to occupy certain sets of bond sites than others. The electrons will tend at all times to occupy the more favorable bond sites for a large fraction of the time.

We now consider the change when one electron spin is reoriented so that there are two more spins of one sign than of the other. These two electrons now require separate orbitals at all times, whereas before they could occupy a single orbital.

Even in a perfect, stationary lattice it seems likely that this change may sometimes cause the use of a less favorable orbital and a finite increase in energy. But in a distorted lattice the new state clearly requires the use of a less favorable orbital, because, in order to maintain uniform charge distribution, the additional orbital must be selected within a very limited region of space. Thus we conclude that unpairing two electron spins will cause a finite increase in energy in a vibrating lattice.

The writer does not claim to have proved rigorously this essential point, but he does believe that the best available theoretical arguments strongly indicate the correctness of this finite excitation energy. Also, he is aware of no disproof.

While we shall not discuss the transition to the normal state in any detail, it seems safe to assume that the transition temperature is of the order of magnitude ϵ/k, where ϵ is the energy required to unpair the first spin. We shall give some explicit support for this relationship later. For a rigid perfect lattice we believe that this energy may still have a nonzero value ϵ_∞ in at least some substances. In all cases, however, we expect the electronic energy to rise quadratically with lattice distortion and to rise slightly more steeply when a spin is unpaired. Thus we expect a term $\epsilon(M)$ proportional to the square of the amplitude of the lattice vibrations or to $M^{-1/2}$ if M is the atomic mass.

Thus we have

$$kT_c \cong \epsilon = \epsilon_\infty + \epsilon(M) = \epsilon_\infty + (\text{const.})\, M^{-1/2}. \qquad (2)$$

If ϵ_∞ is negligible, one expects T_c to vary with $M^{-1/2}$, as was observed for mercury.[6] In the case of tin[7] the reported exponent of M is -0.462 ± 0.014, which deviates from $1/2$ by about three times the stated error. This offers some indication that ϵ_∞ may be significant for tin.

Thus the isotope effect is satisfactorily explained in this theory. It is also interesting to note that these energy terms are very small in comparison with the total zero-point energy of lattice vibration. Consequently no large difference in lattice constant, elastic constants, or lattice specific heat is expected between normal and superconductive states.

II

Next let us consider the possibilities of low-energy thermal excitation of the completely paired electronic system. Pines and Bohm[8] have shown that density oscillations of the usual phonon type have a very high energy in fluid of electrically charged particles and that they will not be excited in the electronic system of metals at low temperatures. The phonons in a metal are to be regarded as lattice motions. The accompanying electronic adjustment for lattice phonon motion is exactly similar to that for lattice zero-point oscillations. The electronic system is still in its ground state.

We have already mentioned the possibility of translational motion of the entire electronic system with respect to the lattice. This requires, of course, that the superconductor be appropriately connected for current flow. There is also the possibility of nonuniform flow, provided the flow rate varies only slowly.

Since all density oscillations are of high frequency and high energy, we may conclude that, as long as all electrons are paired, the only low-energy motion is the translational flow just mentioned. This is such an important point, however, that we shall attempt to show more explicitly the absence of other low-energy states.

By complete electron pairing we mean, of course, that there are equal numbers of electrons with positive and with negative spin. Moreover, we expect that the wave function will be large only for electron locations which constitute an essentially uniform distribution of electrons of each spin as well as uniform distribution of total electronic charge.

Now let us suppose that an electron of positive spin moves slowly from a region A to another region B. In order to maintain uniform charge density (in the absence of net flow), another electron must be transferred from B to A. If the second electron also has positive spin, we do not have an experimentally distinguishable state. We have merely the exchange of two electrons, which leaves the wave function unchanged except for a reversal of sign. This is just the property of the ground state; we have no new state.

The other possibility is that the charge compensation occurs through the motion from B to A of an electron of negative spin. Now we have in region B an excess of two positive spins and in region A an excess of two negative spins. Provided that regions A and B are sufficiently far apart, this is clearly a new state. Its wave function can be orthogonal to that of the ground state. But its energy will exceed that of the normal state by approximately twice the energy of spin unpairing together with any excess electronic kinetic energy. Consequently it is not a low-energy state in the sense of having lower energy than that required to unpair spins.

It is difficult to predict whether there will be states of the type just described where the energy is not sufficient to "dissociate" the regions A and B of excess positive and negative spin. However, detailed consideration indicates that the energies of such states almost certainly must exceed that of a state with a single pair of electrons of parallel spin. Consequently this question is of secondary importance.

We now conclude that in the first approximation the low-energy excited electronic states (other than those for current flow) may be described in terms of unpaired electron spins. If the spin of one electron is reversed, then in the region immediately around it there will be an excess of two electrons of a given spin. These two electrons may then separate to form two regions with single excess spin. Each region may move through the lattice. Thus the location and motion of each region of excess spin comprise observable features which can characterize a large array of excited states.

The momentum associated with such excitations can interact with the lattice phonons of appropriate wave length. Thus the electronic specific heat of the superconductor will be in thermal equilibrium with the lattice. On the other hand, there does not appear to be any mechanism whereby lattice oscillations can directly create such electron-spin excitations. They must be created at a surface or an interior defect or may be transferred into a superconductive region from a nonsuperconductive region of the metal. This picture is in accord, in general features at least, with the observed facts on the kinetics of the phase transition in superconductivity.[9]

The superconductive properties arise because the infinitesimal energy quanta of translation of the entire electronic system are unable to excite either phonons or electron excitations of the spin-unpairing type. For electronic motion to excite phonons the force exerted on the lattice atoms must vary from atom to atom with

appropriate wave length. This flow of the entire electronic system will exert a uniform nonoscillating force and hence cannot excite phonons. In order to create new electron-spin excitations, a larger energy is required than a single quantum of translation can provide, unless the velocity of the general motion is high. The final possible source of resistance is a transfer of energy of general electron motion to the existing electron-spin excitations. Here, as in the case of the phonons, the force on the excitation must vary in space, whereas the general electron motion yields only a uniform force. These energy-transfer prohibitions are similar to those operating for the superflow in liquid He II.

We conclude that there is no mechanism for transfer of the energy of a slow motion of the entire electronic system. Thus we have the infinite conductivity without any conflict with the Bloch theorem, which requires the state with current flow to have higher energy than the state of zero current.

We have already noted the proof by Bardeen[2] that an energy-level system of this type will show the Meissner effect.

Recently Corak and collaborators[10] showed that the electronic heat capacity of superconductors at temperatures below follows T_c the exponential equation

$$C = \gamma T_c a e^{-bT_c/T}, \tag{3}$$

where the constants a and b have values near 9.17 and 1.50, respectively, for several metals. The energy-level pattern given by our theory yields just this type of equation with $\epsilon = bkT_c$. This is evidence for our earlier assumption that the transition temperature is proportional to the energy of excitation ϵ.

Since our energy term $\epsilon \cong kT_c$ depends primarily on conduction electron–lattice interactions which are the cause of resistance in the normal metal, we may conclude that T_c will be roughly proportional to the resistance for metals with equal densities of conduction electrons. While superconductors generally have high resistance in the normal state, one cannot claim that any such quantitative relationship holds. The variety of complicating conditions, however, is such that lack of quantitative agreement is not surprising.

We should note that, in contrast to the proposals of Frohlich and Bardeen,[11] our theory gives no minimum magnitude of lattice interaction below which the metal will fail to become superconducting even at 0° K. We predict merely that T_c will be much lower for metals with small lattice interaction. However, there is another source of interference with superconductivity which will prevent some metals from attaining that state even at absolute zero. This is atomic magnetic moment from inner electron shells or possibly even nuclear magnetic moment. If the effective magnetic field from this source exceeds the critical field for superconductivity, then no superconductivity occurs. This effect eliminates ferromagnetic (also probably antiferromagnetic) metals from the list of possible superconductors. Since the critical field H_0 is roughly proportional to ϵ (and to T_c), even the nuclear magnetic moment might prevent superconductivity in cases where ϵ is very small. However, tests for superconductivity have not been pushed to low enough temperatures and external fields to make the nuclear moment an important criterion yet.

We believe that we have here the basis for a theory of superconductivity. It yields the observed properties of the superconductive state near 0° K. At a later

date we hope to discuss the nature of the transition to the normal state and, if possible, to add some more quantitative aspects to the general theory.

The writer wishes to thank the Guggenheim Foundation for a fellowship and Oxford University for its hospitality during the period this paper was being prepared.

[1] M. R. Schafroth, *Phys. Rev.*, **96**, 1149, 1954; **100**, 463, 1955.

[2] J. Bardeen, *Phys. Rev.*, **97**, 1724, 1955.

[3] See the treatment of Feynman, in C. J. Gorter (ed.), *Progress in Low Temperature Physics* (Amsterdam: North Holland Publishing Co., 1955), chap. ii.

[4] See, for example, R. E. Peierls, *Quantum Theory of Solids* (London: Oxford University Press, 1955), p. 108.

[5] L. Pauling, *Nature of the Chemical Bond* (2d ed.; Ithaca, N.Y.: Cornell University Press, 1945), p. 402; *Proc. Roy. Soc. London, A*, **196**, 343, 1949.

[6] C. A. Reynolds, B. Serin, and L. B. Nesbitt, *Phys. Rev.*, **84**, 691, 1951.

[7] J. M. Lock, A. B. Pippard, and D. Shoenberg, *Proc. Cambridge Phil. Soc.*, **47**, 811, 1951.

[8] D. Pines, *Phys. Rev.*, **92**, 626, 1953; see also D. Bohm and D. Pines, *Phys. Rev.*, **85**, 338, 1952; **92**, 609, 1953.

[9] See T. E. Faber and A. B. Pippard, in C. J. Gorter (ed.), *Progress in Low Temperature Physics* (Amsterdam: North Holland Publishing Co., 1955), chap. ix.

[10] W. S. Corak, B. B. Goodmann, C. B. Satterthwaite, and A. Wexler, *Phys. Rev.*, **102**, 656, 662, 1956.

[11] J. Bardeen, *Revs. Mod. Phys.*, **23**, 261, 1951, and references there cited, including H. Fröhlich, *Phys. Rev.*, **79**, 845, 1950.

---

*CONVERGENT APPROACH TO THE THEORY OF ELEMENTARY PARTICLE INTERACTIONS**

By I. E. Segal

UNIVERSITY OF CHICAGO

Communicated by S. Chandrasekkar, July 13, 1956

1. We sketch here an abstract theory whose results are as yet only qualitative but which is nevertheless encouraging in view of the present state of knowledge about elementary particles. It appears that a logically complete and fundamentally simple theory may be based on the formulation of elementary particles essentially as group representations and that mathematical divergences may be eliminated simultaneously with physically nonoperational features.

2. More specifically, we first indicate how a square-integrable invariant (tensor) of the representations defining a set of elementary particles determines, by means of the duality transform,[1] a convergent S-operator for their coupled fields. This establishes the existence of nontrivial interactions satisfying the covariance criterion of an earlier paper,[2] but in the realistically crucial case of a bounded invariant our procedure is applicable only in a formal sense. It develops, in fact, that in general (in particular for linearly coupled bosons with infinitely many states) there can exist no actual operator on Hilbert space having the formal (invariance and functional dependence) properties appropriate to the S-operator. We therefore turn to a mathematically modified, although symbolically equivalent, formulation and

indicate how a satisfactory theory may be determined by an invariant of the generally relevant type.

The latter approach exhibits in convergent form typical features of "subtraction" physics. Only those self-adjoint operators A satisfying a certain limit condition allowing a meaning to be given to $S^{-1}AS$ correspond to experimental observables. This condition excludes in general the total number or momentum of all particles of a certain type but includes in significant cases occupation numbers and the momentum of particles in individual states.

On the other hand, standard relativistic radiation theory cannot be made completely convergent in this fashion, because the trilinear invariant of the unitary representations associated with the Maxwell photon and the Dirac electron is only infinitesimal, within the framework of abstract representation theory. As a result, these representations with the usual interaction can be formulated within a closed system only if a continuum of other kinds of elementary particles is included. It seems more likely for a variety of reasons[3] that what are really involved are slightly modified representations of a slightly different group admitting a proper invariant.

3. It is no essential loss of generality to treat the interaction of several, or even continuously many, fields as a coupled boson-fermion field, all the elementary bosons (resp., fermions) being combined into a "universal" unitary representation U_b on (the Hilbert space) H_b (resp., U_f on H_f), whose spectrum[4] consists of the elementary particles in question. We assume given (mathematical) conjugations J_b (resp., J_f) that commute with the respective representation operators and play the role of fundamental symmetries.[5] The state space of the field then consists of the Hilbert space $K = K_b \otimes K_f$, where K_b denotes the Hilbert space of all square-integrable covariant symmetric tensors over H_b, and K_f is correspondingly related to skew-symmetric tensors over H_f. Now let w be a tensor in K that is invariant under the induced action $a \to \Gamma(a)$ of the basic group G as well as the induced action of the fundamental conjugation. The corresponding S-operator is defined as $S = D^{-1} \exp(iL_{Dw})D$, where D is the direct product of the duality transforms, from K_b and K_f to the spaces of square-integrable functionals over the real subspaces of H_b and H_f left fixed by the conjugations J_b and J_f, with respect to the canonical normal and Clifford distributions, respectively,[1] while L_u denotes the operation of left multiplication by u. The observable A' of the interacting fields corresponding to the free-field observable A is then defined as $S' = S^{-1}AS$. The invariance properties of w are reflected in the commutativity of S with the $\Gamma(a)$ and in its unitary character. It is immediate that the theory is covariant: $[\Gamma(a)^{-1} A\Gamma(a)]' = \Gamma(a)^{-1} A'\Gamma(a)$.

4. There is an essentially well-known mutual correspondence between invariants of representations and special structure of their Kronecker products, but in the case of infinite-dimensional representations the tensor that arises from such a special structure need not be "square-integrable" but only "bounded." This is associated with the nonexistence of a true S-operator for the corresponding fields.

The relevant convergence questions are well represented by the case of linearly coupled bosons defined by equivalent representations.[6] Let V be the representation on the space M defining one of the bosons, and assume that V is irreducible and neutral. It is convenient to use the frame of reference provided by the duality

transform, according to which the state space of the coupled fields may be taken as the space $L_2(M' \oplus M', n)$ of functionals $f(x, y)$ of the variables x and y, which vary independently over the real subspace M' of M left fixed by the fundamental conjugation J, that are square-integrable with respect to the isotropic centered normal distribution n. The induced action of the group is then $f(x, y) \to f(V(a)^{-1}x, V(a)^{-1}y)$, where a is arbitrary in G, and the induced action of the fundamental conjugation is ordinary complex conjugation, $f(x, y) \to \overline{f(x, y)}$. The only bilinear invariant is, as a function on $M \oplus M$, the inner product $u \cdot Jv$, and when M is finite-dimensional, the corresponding S-operator is simply the operation of multiplication by $e^{ix \cdot y}$: $f(x, y) \to e^{ix \cdot y} f(x, y)$. When M is infinite-dimensional, $e^{ix \cdot y}$ remains a well-defined *function* on $M' \oplus M'$, but it is not a *measurable functional* with respect to the normal distribution, and the indicated transformation does not exist as an operator on Hilbert space.

5. The meaningless character of the inner product as a measurable functional (or random variable) arises from the circumstances that x and y are both normally and independently distributed; when either x or y is fixed, $x \cdot y$ is a measurable functional of the other variable. This may be seen by writing $x \cdot y$ in terms of an orthonormal basis as $x_1y_1 + x_2y_2 + \ldots$; here $x_1, x_2, \ldots$ and $y_1, y_2, \ldots$ are two sequences of independently distributed normal random variables of zero mean and unit variance. It is then apparent from the theory of sequences of independent random variables that the indicated series converges in no reasonable analytical sense.

Now this might not be completely conclusive as regards the nonexistence of the S-operator, as conceivably a matter of interpretation is involved. But any interpretation must satisfy the following formal conditions: (a) invariance under the orthogonal group on M', or at least under the subgroup consisting of the representation operators $\Gamma(a)$; (b) functional dependence on the operations of multiplication by the x_j and y_k, and so at least commutativity with all operators commuting with the x_j and y_k. It is a fact that no such operator exists, under relatively general assumptions about the basic group G and the representation V; e.g., it suffices that V be irreducible on a noncompact simple subgroup, such as the Lorentz group.[7]

6. Although there is no S-operator, there is an entirely finite S-matrix, which, however, is neither unitary nor proportional to a unitary matrix. To see this, take as an orthonormal basis for the state space of the field in the duality frame the functionals $f_{m,n}(x, y) = f_{m(1)}(x)\, f_{n(1)}(x)\, f_{m(2)}(x)\, f_{n(2)}(x) \ldots$ describing the state in which there are present $m(j)$ bosons of the first kind with wave function e_j, and $n(j)$ bosons of the second kind with wave function e_j, where $e_1, e_2, \ldots$ is an orthonormal basis for M', and the $m(j)$ and $n(j)$ are non-negative integers of which all but a finite number vanish. Putting S_r for the operation of multiplication by $\exp(i \sum_{k=1}^{r} x_k y_k)$, it is readily verified that the relative transition probability amplitudes $(S_r f_{m,n} \cdot f_{m',n'})/(S_r f_{0,0} \cdot f_{0,0})$ converge to finite values $A_{m,n,m',n'}$ as $r \to \infty$. One readily computes, in fact, $A_{m,n,m',n'} = \prod_{k=1}^{\infty} a_{m(k),n(k),m'(k),n'(k)}$, where the a_{ijkm} are given by the generating function $\sum_{i,j,k,m=1}^{\infty} a_{ijkm} p^i q^j s^k t^m/(i!j!k!m!)^{(1/2)} = \exp[(p^2 + q^2 + s^2 + t^2)/4 - (sp + iqt)/2]$. The use of generating functions shows that

$\sum_{m,n} |A_{m,n,m',n'}|^2$ is divergent in case m' and n' describe the states in which precisely one boson of wave function e_1, of either kind, is present. Thus the S-matrix is not proportional to any unitary matrix,[8] a fact which is also deducible from symmetry considerations.

7. What is primarily significant, however, is the existence of self-adjoint operators on Hilbert space corresponding to the formal expressions $S^{-1}AS$, for operators A corresponding to experimentally crucial observables. In particular, the canonical P's and Q's, the occupation numbers of individual states (or of finite-dimensional manifolds), the momenta of each of a dense set of individual states (or of corresponding finite-dimensional manifolds), are represented by well-defined operators for the coupled fields. Specifically, we may define a bounded operator A on K as a (convergent, possibly complex) *observable* if the strong limit of $S_{F'}{}^{-1}AS_{F'}$ exists, as the finite-dimensional subspace F' of M' increases to M'. The observability of an unbounded self-adjoint operator A is reduced to the case of bounded operators by requiring that $\lim_{F' \to M'} S_{F'}{}^{-1} \exp(itA)S_{F'}$ exist and yield a continuous one-parameter group, $-\infty < t < \infty$.

Consider the canonical field operators P_j and Q_k for the bosons of the first kind. The corresponding operators P_j' and Q_k' for the coupled fields may be defined as follows. The problem of dealing with $\exp(itP_j')$ reduces, by virtue of the independence of $S_{F'}{}^{-1} \exp(itP_j)S_{F'}h$ of F' for large F' and any fixed tame function h, to the proof of the isometry of the action of the limit on tame functions. This, in turn, reduces as in the case of a similar problem in an earlier paper[1] to a straightforward one-dimensional computation. In the case of the Q's there is no change; $Q_k' = Q_k$. The P_j' and Q_k' satisfy the same commutation relations as the P_j and Q_k, but it may be shown that there exists no unitary operator on K that transforms the P_j and Q_k into the P_j' and Q_k', showing in still another way the nonexistence of an appropriate S-operator. An argument similar to the foregoing shows the observability of occupation numbers of individual states.

For any self-adjoint generator R of G, $d\Gamma(R)$, where $d\Gamma$ denotes the infinitesimal representation determined by Γ, may be called the total free-field R-momentum, using "momentum" in a generalized sense to cover any dynamical variable in the Lie algebra defining the group in question. The total free-field R-momentum of the bosons of the first kind with wave functions in F may be defined correspondingly as $d\Gamma(R_F)$, where $R_F = p_F R p_F$ and p_F denotes the operation of projection of M onto F. The observability of this operator, for finite-dimensional F, follows as above by reduction to the action on tame functionals of $x \cdot e_1, x \cdot e_2, \ldots x \cdot e_n$, where F is contained in the complex linear subspace determined by $e_1, e_2, \ldots, e_n$.

On the other hand, the total population of bosons of the first kind, or their total momenta, are not observables in general. To show that the total population is not observable, it suffices to show that, for some t, the transform under $S_{F'}$ of the field action induced by the transformation $x \to e^{it}x$, $y \to y$, does not converge. Taking $t = \pi$, the field action is the Wiener-Plancherel transform with respect to the variable x, and the action on the function that is identically one is readily computed and seen not to converge. Now turning to the case of the total F-momentum of the bosons of the first kind, $S_{F'}{}^{-1}\,\Gamma(\exp(itR))S_{F'}$ takes $f(x, y)$ into the product of $f(V(\exp(itR_F))^{-1}x, y)$ with $\exp[i(x - V(\exp(itR_F))^{-1}x) \cdot y]$, where F

is the complex extension of F'. Now as $F' \to M'$, $V(\exp(itR_F)) \to V(\exp(itR))$ strongly, and the question reduces in part to the measurability of $Tx \cdot y$ as a function of x and y for the bounded normal operator $T = V(\exp(itR)) - I$. This requires that T be completely continuous,[7] so that the total R-momentum is observable only when $V(\exp(itR)) - I$ is a completely continuous operator, $-\infty < t < \infty$, an unrealistically severe restriction and an impossible one if, e.g., G is an open simple Lie group.

8. In the case of higher-order interactions the same general procedure applies, but there tend to be fewer observable operators, and there are greater mathematical difficulties in their determination. In the situation of Section 3, any finite-dimensional real subspaces F_b' and F_f' of H_b and H_f determine as in the case of linearly coupled bosons a square-integrable invariant, which converges to the given bounded invariant as the finite-dimensional subspaces become arbitrarily great. The observable operators are those whose transforms via the unitary operators $S_{F_b', F_f'}$ deduced from the approximating tensors are convergent as $F_b' \to H_b'$ and $F_f' \to H_f'$.

In the important special case of the linear-bilinear coupling between a neutral boson and a charged fermion, a covariant coupling formally similar to that in quantum electrodynamics exists (resp., is unique) if and only if the boson representation is contained discretely (resp., essentially once) in the skew-symmetric Kronecker square of the fermion representation. As is suggested by the divergent character of the usual theory of the coupled photon-electron fields, this condition is violated in the case of the Maxwell photon and the Dirac electron, and, in fact, the Lorentz group cannot be used as a basis for a realistic theory of the present type, as the Kronecker products of its relevant representations decompose continuously.[9] On the other hand, there exist other continuous groups having discrete series of irreducible representations, whose Kronecker products decompose discretely in part and which possess representations approximating in a sense those of the Maxwell photon and the Dirac electron. The present situation in the classification of elementary particles gives an additional indication for a group close to, but distinct from, the usual extension of the Lorentz group, and, in particular, isotopic spin may well be explainable in terms of a new outer automorphism analogous to the usual reversal operations.

9. We mention briefly certain formal matters that are not directly pertinent to convergence questions but are needed in the physical interpretation of the foregoing. The use of the canonical conjugation involves the following further axioms: (1) only those operators that commute with the canonical conjugation may represent observables; (2) the observable physical momenta are not precisely the generators of G but rather the self-adjoint products of such generators with the sign of the generator of "translation in time." These assumptions are consistent with the fundamental requirements of quantum phenomenology.[10] They insure the positivity of the energy and eliminate in a covariant fashion the need for hole theory. As special consequences, the parasitic Lorentz momenta are eliminated as observables, and the central element of order 2 in the universal covering group of the Lorentz group (which should persist in any more exact theory) gives a universal elementary charge operator which clarifies the principle of conservation of charge.

The automorphism group of the basic group must also be represented by transformations on the one-particle space, and this gives rise to some additional elements of structure in the definition of elementary particles, which, however, do not affect quantization procedures. The use of the fundamental conjugation makes it possible to insure that the outer automorphisms, which include the usual reversal operations, are represented by unitary operators rather than in part by anti-unitary ones.

10. It is probably premature to speculate on the structure of the system of all elementary particles, but it should be illuminating to give a simple example of a mathematically conceivable form for it. Let G be an arbitrary continuous group, and take as the measure on G the essentially unique left-translation-invariant one. Let H_b denote the space of all square-integrable skew-symmetric functions $g(x, y) = -g(y, x)$ on $G \otimes G$, and define $U_b(a)\ g(x, y) = g(a^{-1}x, a^{-1}y)$. Now let H_f denote the space of all square-integrable functions on G that transform completely inequivalently to any subrepresentation of U_b under the regular representation $h(x) \to h(a^{-1}x)$, and define $U_f(a)\ h(x) = h(a^{-1}x)$ for h in H_f. The fundamental conjugations J_b and J_f may be defined as ordinary complex conjugation, and representatives $U_b'(\alpha)$ and $U_f'(\alpha)$ for an arbitrary automorphism α of G may be defined as follows: $U_b'(\alpha)\ g(x, y) = g(\alpha^{-1}(x), \alpha^{-1}(y))\mu_\alpha$ and $U_f'(\alpha)\ h(x) = h(\alpha^{-1}(x))\lambda_\alpha$, where μ_α and λ_α are positive constants uniquely determined by the condition that the operators be unitary. Then $\int\int g(x, y)\ h(x)\ h(y)\ dx\ dy$ is a linear-bilinear invariant under the representations U, the fundamental conjugation, and the representations U', and therefore gives rise to a mathematically consistent, well-determined, closed system of interacting elementary particles.

An identification with a physical system depends on the labeling of the mathematical operators in terms of experimental observables, which requires further specification of G. The discrete spectra of U_b and U_f give countable families of elementary particles, in which suitable pairs will form a closed subsystem formally analogous to the photon-electron system. The nontriviality and uniqueness of such couplings (and hence the classification of the elementary particles observable as such) would, however, depend on the specific decomposition of the Kronecker products of the relevant representations of G. At present there is no clear physical analogue for the contribution of the continuous spectra of U_b and U_f or, more relevantly, the continuous parts of the decompositions of the Kronecker products of the representations in the discrete series of elementary particles, but it seems plausible that their effect may in part resemble gravitation. The continuous spectrum may reasonably be expected to be of no greater significance in the treatment of systems of a small, finite number of types of particles than in atomic theory.

* Work done in part under a contract between the Office of Naval Research and the University of Chicago.

[1] *Trans. Am. Math. Soc.*, **81**, 106–134, 1956; *Ann. Math.*, **63**, 160–175, 1956.

[2] These PROCEEDINGS, **41**, 1103–1107, 1955. We point out that the assumptions under which the uniqueness of coupling constants is proved in this note do not in general hold in the present theory, although they are valid in certain standard versions of relativistic quantum field theory.

[3] Cf., in particular, *Duke Math. J.*, **18**, 221–265, 1951, last section.

[4] We define the "spectrum" of a group representation as the collection of its irreducible constituents. This is a generalization of the notion of the spectrum of a self-adjoint operator A,

which is equivalent to the spectrum in the present sense of the one-parameter group $\{e^{itA}; -\infty < t < \infty\}$. We make the analogous distinction between the discrete and the continuous spectra of a representation.

[5] An irreducible representation may be called "neutral" if it commutes with some conjugation, i.e., a transformation C such that $C^2 = I$, $C(\gamma x + y) = \bar{\gamma}x + y$ for any complex scalar γ, and $Cx \cdot Cx = x \cdot x$. The conjugation is then unique within sign, and either one may be taken as fundamental. A charged representation is defined as a non-neutral representation that is unitarily equivalent to its transform under some conjugation. An essentially unique conjugation then arises through the inclusion of the antiparticle. In either case the representation space is irreducible under the combined action of the representation operators and the fundamental conjugation. Particles that are neither charged nor neutral have not as yet been proved to exist physically and may in any case be treated in the same way.

[6] We speak of bosons of the first and second kind. In the present theory a (convergent) bilinear coupling between neutral elementary bosons is possible only when their representations are equivalent, and it is convenient and no loss of generality to assume that the representations are identical. The bilinear self-interaction of a single boson field could be treated along the same lines.

[7] This is proved by the author in a paper submitted to a mathematics journal.

[8] Except in an essentially rhetorical sense, i.e., via an infinite constant. In a somewhat more rigorous sense, it is a transform of a unitary operator by an unbounded operator.

[9] Cf. G. W. Mackey, *Ann. Math.*, **58**, 193–221, 1953.

[10] Cf. *Ann. Math.*, **48**, 930–948, 1947.

---

*TEMPERATURE INDEPENDENCE IN A UNICELLULAR "CLOCK"**

BY VICTOR G. BRUCE† AND COLIN S. PITTENDRIGH

DEPARTMENT OF BIOLOGY, PRINCETON UNIVERSITY

Communicated by Max Delbrück, June 21, 1956

Introduction.—Endogeneous diurnal rhythms have been observed in living organisms of all types, from micro-organisms to mammals, and the endogenous nature of these rhythms has long been recognized. Recent interest in these rhythms has been stimulated by the investigations of Frisch,[1] Kramer,[2] Pardi and Papi,[3, 4] and others which have shown that the celestial navigation of bees, birds, and amphipods is made possible by a biological chronometer of the endogenous type. The implication that endogeneous rhythms serve generally as a time sense for organisms is reinforced by several observations of temperature independence of the period of these rhythms. Clearly, if the rhythm is to be an accurate time-giver in nature, its rate must be effectively independent of temperature within ecological limits. Temperature independence of the clock systems in poikilotherms has been demonstrated as long ago as 1932 by Wahl,[5] using the bee. More recently, Brown and Webb[6] showed temperature independence of the rhythms in the crab *Uca* and the clam *Venus;* Pittendrigh[7] found the rhythm of eclosion in *Drosophila* to be independent of temperature; and Rao[8] showed the same for a tidal rhythm in the mussel *Mytilus.* Recent work on persistent daily rhythms in plants (Ball and Dyke;[9] Bünning and Leinweber[10]) has also shown the same insensitivity of the period to the temperature regime.

In 1948 Pohl[11] showed that the single-celled motile green alga *Euglena gracilis* has an endogenous rhythm, but he did not investigate the temperature character-

istics of this rhythm. The present paper reports temperature independence of the *Euglena* rhythm. Thus the endogenous rhythm in *Euglena* exhibits the important characteristics of the endogenous rhythms in higher organisms without the complexity of multicellular organization.

Nature of the Endogenous Rhythm in Euglena.—The rhythm in *Euglena* which has been investigated is a rhythm in the phototactic response. *Euglena* has an eye spot (concentration of pigment) and is photosensitive, in addition to being motile. The phototactic response is thus dependent on the motility and the photosensitivity, and by the method employed in the present investigation these factors cannot be separated. Although many factors affect the phototactic response, conditions may be standardized sufficiently so that the pattern of phototactic responses obtained in one day is repeatable the next day. In the present investigation a narrow beam of light (intercepting about 2 per cent of the culture volume) is sent through a suspension of algae. The beam of light falls upon a photocell, and the current from the photocell is recorded automatically. This beam of light serves two main functions. It acts as a "light-trap" to attract the *Euglena*, and it serves also as part of the sensing system to measure how many cells have been attracted to the light and how quickly they are attracted there. The beam of light, which will hereafter be called the "test light," is turned on for a short time (generally for 30 minutes) at periodic intervals (usually every 2 hours). The recorded pattern of the reduction in current from the photocell during the exposure of the culture to the test light is a measure of the aggregation of *Euglena* in the beam of light and will be referred to as the "phototactic response." The phototactic response is thus measured in a population of cells and not in individuals. There is a daily rhythmical change in the character of the *Euglena* phototactic response assayed in this way: the cells are positively phototactic by day and show little or no response at night. This is a true persistent daily rhythm in the sense that it endures in the absence of any 24-hour cycle of light or temperature.

Materials and Methods.—*Euglena gracilis*, strain z, obtained from Dr. S. Hutner, of the Haskins Laboratories in New York, has been used for these experiments. The experiments reported here have all been done with *Euglena* grown in simple inorganic medium of the following composition: KNO_3 0.1 per cent, $MgSO_4 \cdot 7H_2O$ 0.01 per cent, K_2HPO_4 0.02 per cent, $FeCl_3$ 0.0001 per cent, $CaCl_2$ 0.0001 per cent. Cells in cultures grown in this medium remain viable and active for months if they receive sufficient light.

All observations are made and recorded automatically, using a system patterned after that used by Pohl in his investigation. The apparatus is shown schematically in Figure 1. Approximately 10 ml. of culture are contained in a 50-mm.-diameter Carrel flask. A microscope lamp directly below the culture throws a vertical beam of light about 7 mm. in diameter (the test light) on to a Silicon Solar Cell. An adjustable fraction of the voltage output of the Solar Cell is recorded without amplification on a Brown Strip Chart Electronik Potentiometer. A so-called "day lamp" consisting of a 4-watt fluorescent lamp is located laterally to the culture and provides illumination for a day-night cycle. In operation the test lamp has generally been turned on for 30 minutes every 2 hours, and the Chart Recorder only operated when the test lamp was on; thus the test records generally consist of half-hour "responses" at 2-hour intervals. The day lamp is always off when the

test lamp is on. The culture vessel is mounted in a black lucite chamber, and the whole assembly is kept in a light-tight, temperature-controlled cabinet.

The nature of the test setup is such that the phototactic response of a culture grown and kept in strictly constant conditions cannot be tested. The only conceivable way in which this could be done would be to remove aliquots periodically from a culture kept in constant conditions and to test the aliquots for phototactic response, after which they could be discarded. This is impractical not only because shaking affects the phototactic response but also because the rhythm is lost in constant light and because the cells become inactive in continuous darkness. Thus the rhythm of phototactic response must be studied under conditions in which the culture is exposed to the repeated stimuli of the test lamp. For convenience, subsequent descriptions will refer to such conditions as "constant conditions."

Pattern of Rhythm in Alternating Light and Darkness (LD).—Figure 2 (April 13) shows a typical record of phototactic responses of a culture which has been exposed to a twice-repeated cycle of 12 hours of the 4-watt fluorescent day lamp followed by 12 hours of darkness with dawn at 10:00 in the morning. This cycle of day-lamp light and dark is referred to as "LD conditions." The test lamp was turned on for 30 minutes at 2-hour intervals on the odd-numbered hours. The detailed nature of the pattern of responses depends on several factors, including the light intensity (and color) of the day lamp, the intensity, duration, and repetition frequency of the testing light, and the temperature. The period of the rhythm is, however, always 24 hours in an LD cycle of 24-hour periodicity.

Pattern of the Rhythm in "Constant Conditions."—With the day lamp continuously on (LL), no continuing rhythm has been observed. This loss of an overt periodicity in conditions of constant light, although not understood, is a feature common to many of the persistent rhythms. No study has been made of the way in which the *Euglena* rhythm is lost after transfer to LL.

With the day lamp continuously off (DD), a rhythm does persist as shown in Figure 2 (April 14 and 15) which is taken from an experiment using 30-minute test lights. Using a shorter duration for the test-light cycle (approximately 10 minutes every 2 hours), however, even the DD rhythm is rapidly lost. Thus the persistence of the phototactic rhythm in DD is evidently dependent on the cells receiving sufficient energy via photosynthesis from the light of the test lamps.

It is necessary to record for at least 4 or 5 days in order to get a reasonably accurate estimate of the period length in DD. It greatly simplifies the graphical representation of the data to record only the potentiometer reading corresponding to the maximum response of the *Euglena* during each 2-hour interval, and this seems to be a quite satisfactory method for determining the average period. Successive days are plotted one below another, with a constant displacement of the potentiometer scale from day to day. Any departure of the period from 24 hours thus shows up as a gradual shifting of the time of day at which the minimum response occurs. The persistence of an overt rhythm in DD with a period which is not exactly 24 hours or an integral multiple of the period of the test lamps is evidence that we are dealing with an endogenous rhythm.

Temperature Independence.—Figures 3 and 4 show the results of DD experiments at different temperatures. It can be seen that the period of the rhythm at 33°,

26°, 23°, and 16.7° is always very close to 24 hours.[12] In this sense the period of the rhythm is considered to be essentially temperature independent. At the lower temperature there is a more noticeable damping of the rhythm than at the other temperatures. This may be due to a more pronounced nonsynchrony of the cells in the culture, or it may be due to the lowered efficiency of conversion of light

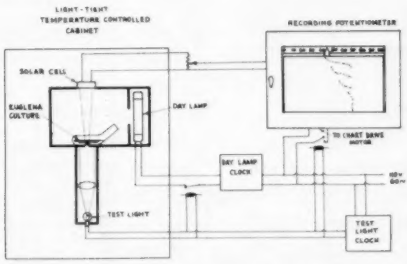


FIG. 1.—Schematic diagram of apparatus for testing and recording the phototactic response in *Euglena*.

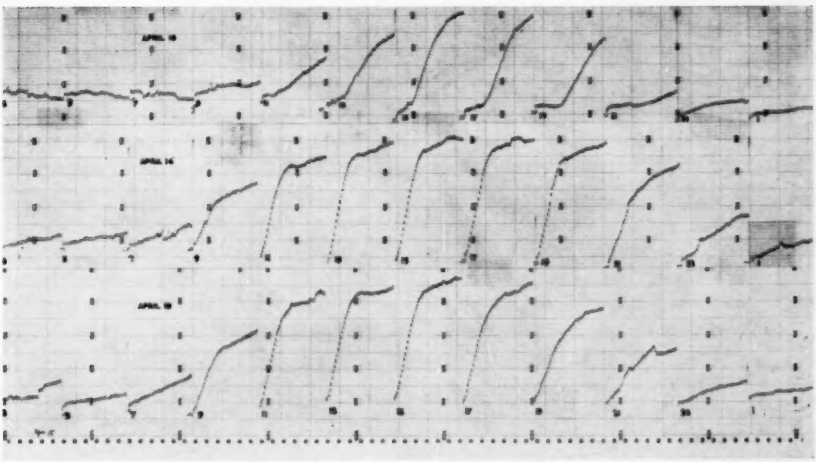


FIG. 2.—Potentiometer chart record of phototactic response. The greater amplitude of the "responses" at the hours 11, 13, . . ., 21 corresponds to a greater change in the voltage across the photocell as a result of the aggregation of the *Euglena* in the test beam. The test lights were on for 30 minutes every 2 hours. April 13, LD: The day lamp was on from hour 10 to hour 22. April 14 and 15, DD: The day lamp was off. Temperature 20° C.

energy at the lower temperature. At all temperatures, however, there is a significant damping of the rhythm, and this renders impossible a precise quantitative characterization of the very small temperature dependence suggested by the data.

Other Similarities of the Euglena *Rhythm to Rhythms in Higher Organisms.*—The endogenous rhythm exhibits characteristics other than temperature independence which are similar to those found in higher organisms. One of these is the

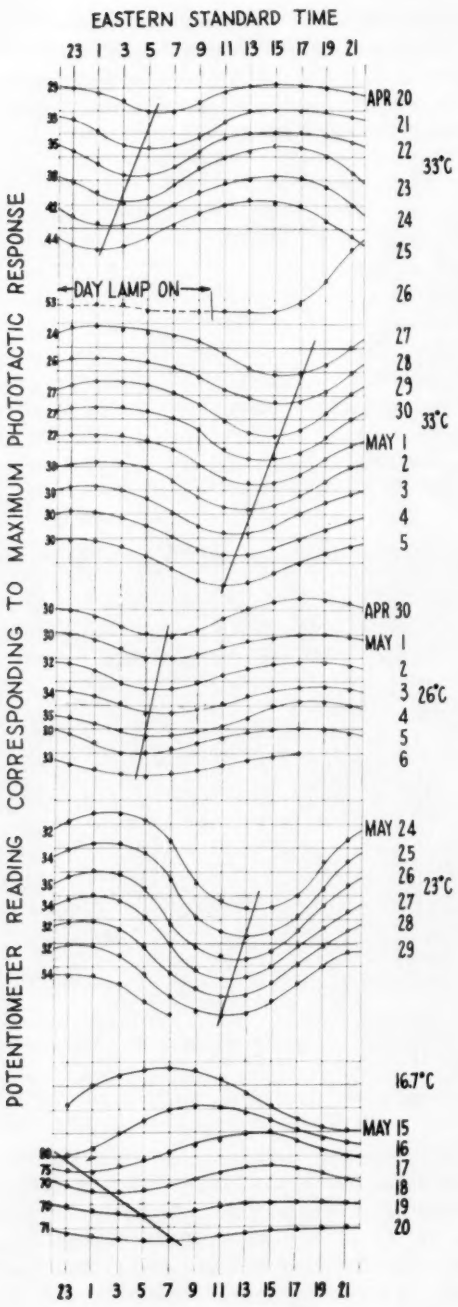


FIG. 3 (left).—Results of experiments illustrating the virtual temperature independence of the diurnal rhythm of phototactic response in *Euglena* in DD conditions. The circled points represent the potentiometer reading corresponding to the maximum phototactic response recorded each time the test lamp is on.

The numbers on the left-hand scale are the potentiometer readings corresponding to hour 22 on successive days. The light horizontal lines are ten units apart, and each day the scale is shifted ten units. The inclined lines passing through the minimum points on the response curves illustrate the slight departure of the period from 24 hours and form the basis for the estimate of the period at the different temperatures. Twelve hours of the day-lamp light was used on May 25–26 to reset the phase of the rhythm as described in the text.

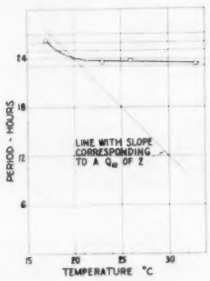


FIG. 4 (above).—Summary of Figure 3, illustrating the temperature independence of the *Euglena* rhythm.

phenomenon of resetting[13] (phase shift) by a single perturbation with light. This is well enough exemplified in Figure 3. The day lamp was turned on for 12 hours on May 25–26. It is clear that this single perturbation establishes an entirely new phase without affecting the period.

Other experiments show that an arhythmic population can be made rhythmic by a single light stimulus. This is precisely analogous to the situation found in *Drosophila*, and, as in the latter case, two possible interpretations exist and cannot be resolved. We are not certain whether a single stimulus initiates rhythms in cells or merely establishes synchrony in a population of rhythms with randomly distributed phases.

Discussion.—No satisfactory physiological explanation of the remarkable fact of temperature-independent endogenous rhythms has been advanced. Three types of mechanisms have been suggested. These are as follows: (1) The diurnal periodicity is maintained by some uncontrolled variable with a diurnal periodicity. (2) The periodicity of the clock derives from some relatively temperature-independent physical (as contrasted with chemical) process. (3) Temperature independence is achieved by virtue of a temperature-compensating mechanism within the organism, although component parts of the "clock" may be temperature dependent. The first mechanism is unsatisfactory for those cases, such as *Euglena*, where the periodicity in "constant conditions" is not exactly 24 hours. The following consideration favors an interpretation in terms of a temperature-compensating mechanism. Repeated and nonrepeated temperature changes have been demonstrated to affect endogenous rhythms in several ways. The entrainability to periodic temperature cycles and the influence of single temperature shocks in initiating rhythms, introducing transients, or shifting the phase of a rhythm are examples which have been demonstrated in several organisms. An argument against the second mechanism hinges on the fact that the diurnal (24-hour) rhythm in *Euglena* persists even when the cells are growing in organic media so rapidly that they divide more frequently than once every 24 hours. There is implied in the second mechanism for achieving temperature independence the idea that the periodicity of the "clock" is intimately tied into the physical localization of the components of the "clock." The ability of the "clock" to continue running while it is duplicating itself is an argument against this mechanism.

No model can be proposed for the clock in *Euglena;* however, we should like to point out that the rhythm possesses certain features which are characteristic of a class of periodic phenomena known in physics and mechanics as self-sustained relaxation oscillations. This aspect of endogenous rhythms has been discussed in some detail by Pittendrigh,[13] and the subject is not sufficiently well explored in *Euglena* to merit further consideration here.

Summary.—The endogenous diurnal rhythm in the phototactic response of *Euglena* has been shown to be temperature independent in the sense that the period of the rhythm at constant temperatures ranging from 16.7° to 33° is always very close to 24 hours. Attention is drawn to the fact that this most important feature of endogenous rhythms has now been demonstrated to occur at the unicellular level and that, in looking for a common mechanism in biological "clocks," one need not go to the complexity of nervous or multicellular organization.

* This work was supported by a grant from the National Science Foundation and by funds from the Eugene Higgins Trust allocated to Princeton University.

† United States Public Health Service Fellow, National Cancer Institute.

[1] K. von Frisch, "Die Sonne als Kompass im Leben der Bienen," *Experientia*, **6**, 210, 1950.

[2] G. Kramer, "Experiments on Bird Orientation," *Ibid*, **94**, 265, 1952.

[3] L. Pardi and F. Papi, "Ricerche sull'orientamento di *Talitrus saltator* (Montagu), (Crustacea-Amphipoda)," *Z. vergleich. Physiol.*, **35**, 459 and 490, 1953.

[4] L. Pardi and M. Grassi, "Experimental Modification of Direction-finding in *Talitrus saltator* (Montagu) and *Talorchestia deshayesei* (Aud.) (Crustacea-Amphipoda)," *Experientia*, **11**, 202, 1955.

[5] O. Wahl, "Neue Untersuchungen über das Zeitgedächtnis der Bienen," *Z. vergleich. Physiol.*, **16**, 529, 1932.

[6] F. A. Brown, Jr. and H. M. Webb, "Temperature Relations of an Endogenous Daily Rhythmicity in the Fiddler Crab, *Uca*," *Physiol. Zoöl.*, **21**, 371, 1948.

[7] C. S. Pittendrigh, "On Temperature Independence in the Clock System Controlling Emergence Time in *Drosophila*," these PROCEEDINGS, **40**, 1018, 1954.

[8] K. P. Rao, "Tidal Rhythmicity of Rate of Water Propulsion in *Mytilus* and Its Modifiability by Transplantation," *Biol. Bull.* **106**, 353, 1954.

[9] N. G. Ball and Inga J. Dyke, "An Endogenous 24-Hour Rhythm in the Growth Rate of the *Avena* Coleoptile," *J. Exptl. Botany*, **5**, 421, 1954.

[10] E. Bünning and F. J. Leinweber, "Die Korrektion des Temperaturfehlers der endogenen Tagesrhythmik," *Naturwissenschaften*, **43**, 42, 1956.

[11] R. Pohl, "Tagesrhythmus im phototaktischen Verhalten der *Euglena gracilis*," *Z. Naturforsch.*, **3b**, 367, 1948.

[12] In Figure 2 the temperature was 20° and, although only 2 days were recorded in DD, the period is obviously very close to 24 hours. Another experiment indicated a period of 25 hours at 18.5° C.

[13] C. S. Pittendrigh, "Perspectives in the Study of Biological Clocks," in *Symposium on Perspectives in Marine Biology* (Berkeley: University of California Press, 1957) (in press).

---

*INFLUENCE OF PROTEIN AND CALCIUM ADDITIONS TO ALREADY ADEQUATE DIET**

BY HENRY C. SHERMAN,† CONSTANCE S. PEARSON, MARY E. R. BAL, ANDREA MCCARTHY, AND CAROLINE SHERMAN LANFORD

DEPARTMENT OF CHEMISTRY, COLUMBIA UNIVERSITY, NEW YORK 27, NEW YORK

Communicated June 25, 1956

Laboratory-bred rats have, in repeated instances, thrived for seventy generations or more upon a uniform diet consisting exclusively of one-sixth dried whole milk and five-sixths ground whole wheat with table salt and distilled water (Diet A, Laboratory No. 16). We emphasize "exclusively" because some readers have mistakenly supposed that these animals received other food occasionally. Perhaps it should also be emphasized that these rat families have continued to thrive until killed for lack of space to keep them longer under the strict laboratory control of these experiments. Families of them are still (1956) thriving in the eighty-second generation—probably uniquely rigorous evidence that this basal diet meets all of these animals' needs generation after generation indefinitely.[1]

This basal Diet A (Laboratory No. 16) is therefore certainly adequate in the strictest sense of any ordinary use of the word; yet, as our previous papers have

shown, it is not optimal but can be improved by increasing its content of calcium or vitamin A or such a natural food as milk. The basal diet contains about 0.2 per cent of calcium and about 14 per cent of protein in the air-dry mixture.

The present paper summarizes those of our experiments in which the basal diet just described has been supplemented by extra protein either alone or with simultaneous increase of calcium content.

Long-Term Nutritional Experiments with Added Protein.—In the first series of these experiments the above-described basal diet (of which the animals ate ad libitum) was supplemented by the addition of 5 gm. of lean meat per capita on 6 days of each week. This was equivalent to increasing the protein content of the diet from about 14 to about 24 per cent of the total solids of the supplemented diet. The protein-supplemented diet is designated "Diet 16P5." Some of the findings of this first series of experiments have already been published.[2]

Such raising of the protein-intake level results, as in the similar experiments of Slonaker,[3] in more rapid growth and earlier puberty. The adult body weight is greater, in both sexes, and the young are heavier at 28 days of age, when the diet contains added protein.

The data on such of the animals as had completed their life-cycles at the time the earlier report was published indicated no effect on the longevity of the males and suggested that the added protein might shorten both the life-span and the "period of the prime" of the females. When account is taken of the animals still living when the 1950 report[2] was written, the final averages again strongly suggest that the addition of protein alone to Diet 16 shortens the length of life and the duration of reproductive activity of the females (Table 1). For the males, on the other hand, raising the protein level had, if anything, a slightly beneficial effect as measured by the length of life.

TABLE 1

FINAL RESULTS ON ANIMALS SHOWN IN TABLE 2 OF 1950 REPORT

	Diet 16	Diet 16P5	Difference with Its PE.	Critica Ratio
Length of reproductive life (days)	394 ± 8.5	315 ± 16	79 ± 18.0	4.4
Length of life (days):				
Males	660 ± 20	701 ± 17	41 ± 26.2	1.6
Females	855 ± 16	770 ± 22	85 ± 28.0	3.0

As was noted in the earlier papers, the life-history of the females after attaining maturity is uncertain. Some of the animals receiving the extra protein maintain higher records than their controls throughout, while others develop nervous instability with symptoms suggestive of calcium deficiency. Analysis of typical animals showed that such increase of the protein level from 14 to 24 per cent of the total solids of the food did result in a lower calcium content of their bodies than in those of their controls.[4] Most of the animals recovered spontaneously, but an appreciable percentage of the females died, often after miscarriage. Their abnormality was undoubtedly due to their diets and might be explained in any of three ways: shortage of calcium at a time of rapid growth, or an imbalance between the intake levels of calcium and protein, or the paucity of the calcium intake may become manifest when the growth rate is accentuated by increased protein intake.

Or, in other words, the lowering of body calcium by protein enrichment of a low-calcium (minimal adequate) diet may be regarded as an effect of a protein-calcium imbalance or of an undue acceleration of growth by the increased protein intake or of both.

The extent of this hazard to breeding females may be illustrated by the following comparison: Of forty females on the basal diet containing 0.2 per cent of calcium, three animals (7.5 per cent) died between the ages of one month ("end of infancy") and one year (early adulthood), while, of forty parallel cases of the same heredity and nutritional background, but receiving the supplement of lean meat as described above, seven animals (17.5 per cent) died during the same age period. Thus in these experiments the death rate of females in this age period was more than twice as high on the higher protein-intake level. Here the imbalance followed increase of the protein, without change in the calcium content, of the diet. Experiments were then made to determine whether this hazard could be reduced or obviated by increasing the intake level of calcium at the same time as that of protein. Qualitatively, this has been found to be the case; and experiments designed to study quantitatively the effects of simultaneous supplementation of the diet with protein and calcium are reported in the next section.

Effects of Simultaneous Enrichment with Protein and with Calcium.—In the second series of experiments three diets were compared: Diet 16P5, the protein-enriched diet of the earlier series, which had a protein level of about 24 per cent and a calcium level of about 0.2 per cent, on the dry basis; Diet 162P5, in which the protein level was kept at 24 per cent but the calcium level was raised to 0.4 per cent; and Diet 168P5, with the same protein level but still higher calcium (0.6 per cent).

Using representative 28-day-old animals from our Diet 16 colony, and applying strictly the principle of littermate control, six breeding lots (each consisting of two males and three females) were started on each of the three diets.[5] In addition, one breeding lot of second-generation Diet 162P5 and two of second-generation Diet 168P5 animals were observed.

In Table 2, summarizing the effects of the three diets, the three right-hand columns represent the average results for those cases only in which littermate parallels were observed, while the three left-hand columns include in the average all reasonably comparable cases observed on the three diets.

In Table 3 the comparison is between those animals which received only the meat supplement and all those which received supplementary calcium as well (whether to the 0.4 or the 0.6 per cent level).

When calcium as well as protein is added to basal Diet A, the growth rates of both males and females are still further increased, and greater body weights are maintained throughout adulthood. These effects are clearly significant from a statistical point of view.

The females given additional calcium reach adulthood at a significantly earlier age and, on the average, maintain the ability to bear young much longer (though the latter effect is not proved conclusively by the number of data at hand). The number of young borne, the number reared, and the average weight of the young at 28 days of age are slightly greater for the calcium–protein-supplemented than for the protein-supplemented groups, but the differences are not statistically significant.

The data suggest (though they do not prove) that the supplementary calcium

may increase the average length of life. As was the situation in a much earlier study[6] in which calcium was added to Diet 16, the benefit appears somewhat greater, and the effect more convincing, in the case of the males than in the case of breeding females.

TABLE 2

EFFECTS OF SUPPLEMENTED WHEAT-AND-MILK DIETS CONTAINING 24 PER CENT PROTEIN AND 0.2, 0.4, OR 0.6 PER CENT CALCIUM

(Mean with Its Probable Error)

	All Comparable Cases			Littermate Parallels Only		
	Diet 16P5 (0.2% Ca)	Diet 162P5 (0.4% Ca)	Diet 168P5 (0.6% Ca)	Diet 16P5 (0.2% Ca)	Diet 162P5 (0.4% Ca)	Diet 168P5 (0.6% Ca)
Males:						
Gain in weight, 28–56 days of age (gm.)	84.3 ± 2.4 (14)*	95.2 ± 2.7 (14)	97.2 ± 2.1 (16)	83.4 ± 2.8 (12)	96.8 ± 2.9 (12)	98.3 ± 2.3 (12)
Weight at 100 days (gm.)	214 ± 3.7 (16)	253 ± 3.0 (17)	248 ± 3.5 (18)	....	....	....
Weight at 210 days (gm.)	293 ± 4.3 (16)	325 ± 3.6 (17)	324 ± 4.5 (18)	....	....	....
Weight at 365 days (gm.)	319 ± 5.2 (16)	349 ± 6.5 (15)	354 ± 4.5 (18)	....	....	....
Weight at 500 days (gm.)	308 ± 8.2 (14)	351 ± 7.0 (14)	341 ± 5.2 (18)	....	....	....
Length of life (days)	667 ± 23 (14)	729 ± 18 (14)	727 ± 20 (16)	642 ± 24 (12)	725 ± 22 (12)	744 ± 24 (12)
Females:						
Gain in weight, 28–56 days of age (gm.)	62.7 ± 1.4 (21)	70.9 ± 1.3 (20)	72.8 ± 1.0 (24)	61.9 ± 1.7 (17)	72.3 ± 1.4 (17)	74.9 ± 1.1 (17)
Weight at 100 days (gm.)	162 ± 2.3 (16)†	177 ± 1.6 (8)	180 ± 2.1 (21)	....	....	....
Weight at 210 days (gm.)	204 ± 2.7 (11)	214 ± 2.3 (12)	211 ± 2.9 (18)	....	....	....
Weight at 365 days (gm.)	220 ± 2.5 (13)	242 ± 5.3 (8)	243 ± 3.2 (15)	....	....	....
Weight at 500 days (gm.)	228 ± 3.9 (15)	246 ± 4.0 (16)	245 ± 2.9 (20)	....	....	....
Age at birth of first young (days)	113.0 ± 3.3 (21)	98.0 ± 1.2 (20)	97.8 ± 2.2 (24)	116.0 ± 4.0 (17)	96.7 ± 1.4 (17)	97.8 ± 2.6 (17)
Number of young borne	41.0 ± 2.6 (18)	46.6 ± 2.5 (20)	44.0 ± 1.7 (23)	40.8 ± 3.0 (14)	39.9 ± 2.2 (14)	42.7 ± 3.1 (14)
Number of young reared	34.7 ± 2.4 (18)	37.0 ± 2.0 (20)	38.3 ± 1.7 (23)	33.5 ± 2.6 (14)	35.3 ± 2.0 (14)	38.1 ± 2.8 (14)
Average weight of young at 28 days of age (gm.)	45.4 ± 0.5	47.2 ± 0.6	44.8 ± 0.5	....	....	....
Length of reproductive life (days)	319 ± 17 (18)	386 ± 16 (20)	351 ± 16 (22)	309 ± 26 (14)	353 ± 14 (14)	342 ± 27 (14)
Length of life (days)	807 ± 19 (18)	870 ± 26 (20)	861 ± 17 (23)	790 ± 21 (14)	873 ± 31 (14)	825 ± 26 (14)

* Figure in parentheses indicates number of cases.

† Numbers vary due to exclusion of pregnant females.

Above and beyond the benefits of calcium supplementation which can be measured objectively, the experienced observers who repeatedly handled the animals remarked further on the improvement which "is visible in their alertness and general appearance and palpable in their superior muscle tone and firmness of subcutaneous tissue."[7]

As between the two higher levels of calcium studied, no consistent or significant differences in response could be observed, either in the appearance of the animals themselves or in their vital statistics.

The Problem of Optimal Intakes.—The present findings take on added interest in the light of earlier experiments in which the calcium level only was increased.[6, 8]

Starting with the basal Diet A, we now see that an increase *either* in the protein level (from 14 to 24 per cent) *or* in the calcium level (from 0.2 to 0.4 or 0.6 per cent) increases the rate of early gain as well as the weight of the offspring at the end of infancy, results in larger adult size, hastens the attainment of maturity, increases the number of young borne and the number reared, and extends the length of life of the males. The influence of added protein is somewhat more marked than the influence of added calcium as regards body weight and gain and somewhat less marked so far as the other effects are concerned. Adding *both* protein and calcium gives better growth of both sexes, and, apparently, longer life for the males, than adding either factor alone.

TABLE 3

EFFECTS OF ADDING CALCIUM TO DIET CONTAINING 24 PER CENT PROTEIN AND 0.20 PER CENT CALCIUM: STATISTICAL SIGNIFICANCE OF DIFFERENCES

	Without Ca Suppl.*	With Ca Suppl.†	Difference with Its P.E.	Critica Ratio
Body weight (gm.):				
Gain, 28–56 days of age:				
Males	84.3	96.3	12.0 ± 3.00	4.0
Females	62.7	72.0	9.3 ± 1.61	5.8
Weight at 100 days of age:				
Males	214	250	36 ± 4.4	8.2
Females	162	179	17 ± 2.8	6.1
Weight at 210 days of age:				
Males	293	325	32 ± 5.2	6.2
Females	204	212	8 ± 3.4	2.3
Weight at 365 days of age:				
Males	319	352	33 ± 6.5	5.1
Females	220	243	23 ± 3.7	6.2
Weight at 500 days of age:				
Males	308	346	38 ± 9.2	4.1
Females	228	245	17 ± 4.6	3.7
Length of life (days):				
Males	667	728	61 ± 27.1	2.3
Females	807	843	36 ± 24.2	1.5
Reproductive success of females:				
Age at birth of first young (days)	113.0	97.9	15.1 ± 3.65	4.1
Length of reproductive life (days)	319	367	48 ± 20.2	2.4
Number of young borne	41.0	45.3	4.3 ± 3.0	1.4
Number of young reared	34.7	37.7	3.0 ± 2.73	1.1
Average weight of young at 28 days (gm.)	45.4	46.0	0.6 ± 0.86	0.7

* Animals on Diet 16P5; see Table 2.

† Includes animals on Diets 162P5 and 168P5; see Table 2.

On the other hand, increase in the calcium intake alone extends the period of the prime (as measured by the duration of reproductive capacity of the females) and seemingly also the length of life of the females, whereas increase of the protein intake alone appears to shorten both the prime of life and the length of life of the females. The question whether females on the high-calcium, high-protein diet either live longer or retain the ability to reproduce longer than females on the basal diet alone appears doubtful from the data at hand. A confident answer would require comparison of littermates on the two diets, which has not been made.

Thus it would seem that increasing only the protein intake of our Diet A (from 14 to 24 per cent) has effects on the life-history which, while beneficial at some points, are nonbeneficial at other points. In order to overcome the potential hazards of the higher protein intake, the calcium intake must also be liberal. There seems, however, no reason to believe that the level of calcium which gives the best

results when the protein intake is 24 per cent is any higher than the level of calcium which gives the best results when the protein intake is 14 per cent. The present studies seem to indicate that the most beneficial level of calcium intake when the protein intake is liberal lies in the range of 0.4–0.6 per cent, or two to three times the level of minimal adequacy. In earlier studies,[8] where calcium alone was added to Diet A, the level of optimum intake appeared to be at least three, and perhaps as much as four, times the level of minimal adequacy.

Finally, it would seem justifiable to conclude that a diet liberal both in calcium and in protein is clearly superior to a diet liberal only in protein.

The authors wish to express their deep appreciation to Dr. Charles Glen King for having made possible the completion of this study. In particular, the junior authors wish to thank him for his continued support, encouragement, and guidance after it became necessary for the senior author to withdraw from active participation in the research.

* This work has been aided by grants to Columbia University, at first from the John and Mary Markle Foundation, more recently from the Nutrition Foundation, Inc., the Williams-Waterman Fund in the Research Corporation, and Swift and Company.

† Deceased October 7, 1955.

[1] A comparison will soon be published of the life-histories of such late-generation Diet A animals with those of their ancestors whose families had been on the diet only about half as many generations and with those of their early forebears whose families had only recently been transferred to the diet.

[2] H. C. Sherman, C. S. Pearson, and M. E. Bal, these PROCEEDINGS, **36**, 106–109, 1950.

[3] J. R. Slonaker, *Stanford Univ. Pubs., Univ. Ser., Biol. Sci.*, Vol. **6**, No. 4, 1939.

[4] H. C. Sherman, M. S. Ragan, and M. E. Bal, these PROCEEDINGS, **33**, 356–358, 1947.

[5] A heat wave of extraordinary severity resulted in the untimely deaths of several of the females, reducing to 17 the number of strictly parallel trios observed until maturity and to 14 the number of strictly parallel trios of females observed through the entire normal life-span.

[6] H. C. Sherman and H. L. Campbell, *J. Nutrition*, **14**, 609–620, 1937.

[7] Henry C. Sherman, in report to Williams-Waterman Fund for 1951–1952.

[8] F. O. Van Duyne, C. S. Lanford, E. W. Toepfer, and H. C. Sherman, *J. Nutrition*, **21**, 221–224, 1941.

---

*MEASUREMENT OF IMPOSED VOLTAGE GRADIENT ADEQUATE TO MODULATE NEURONAL FIRING**

BY C. A. TERZUOLO AND T. H. BULLOCK

DEPARTMENT OF ZOÖLOGY, UNIVERSITY OF CALIFORNIA, LOS ANGELES

Communicated by H. W. Magoun, July 6, 1956

Many authors[1–31] have described the effects of polarization by imposed electric current upon nerve cells. We have not seen in the literature, however, a quantitative evaluation of the sensitivity of nerve cells to electric fields in terms of voltage gradient across some appropriate dimension of the neuron. We have undertaken to estimate the threshold value as being the unique value of greatest interest and have found this to be far lower for modulation of the frequency of an already active neuron than for the excitation of a silent one.

In selecting the measurement to be preferred for evaluating such an electric field, the estimation of voltage drop across some chosen region is a second choice to the estimation of current density through a known strategic membrane; but since we do not know how to measure the latter, we are forced to this choice. Furthermore, the selection of structures across which to measure voltage drop presents difficulties. At first glance it may appear that the structure of choice would be the cell membrane and that the inside-outside potential should be the most significant one to measure before and during threshold polarization. Upon further consideration, however, it will be realized that there will be no characteristic value for this membrane potential change, since in an equatorial region of the cell, with respect to the axis of polarization, the potential across the membrane will not be changed at all during polarization, and on one side of this line it will be increased and on the other side decreased. It will be shown, moreover, that the maximal values to be expected near the poles will be virtually beyond the limits of the sensitivity of the high impedance electrode wide-band amplifier systems we have used.

As an approximation, therefore, since we must measure the potential between two points rather than between the whole dendritic surface and the axon hillock, for example, we have measured the gradient across the whole soma in the external field, in an axis of polarization shown experimentally to be the most effective.

MATERIALS AND METHODS

The requirements for a suitable preparation for the above measurement are as follows. An accessible neuron, whose activity can be followed continuously with assurance that this activity originates in the cell under measurement, must be active at a steady level, so that alteration of its frequency by the experimental condition will be noticeable. In addition, we must be able to exclude the possibility that the effects obtained are exerted through presynaptic alterations in other neurons. An ideal preparation which satisfies these conditions is the nonadapting stretch receptor of Crustacea, described by Alexandrowicz[32] and studied physiologically by Wiersma, Furshpan, and Florey,[33] Kuffler,[34] and Eyzaguirre and Kuffler.[35–37] This is a sensory neuron whose cell body lies in the periphery, close to a special muscle fiber bundle (Florey and Florey[38]) to which it is connected by several dendrites emerging from one side, while from the other side of the soma an axon takes origin to run to the central nervous system. Under steady stretch, this cell maintains a steady discharge of impulses at a certain frequency, and this discharge is highly rhythmic, so that very small alterations in its frequency are conspicuous.

We have used the abdominal receptors in the crayfish (*Cambarus clarkii*). These were prepared under a microscope, and the muscle bundle of the nonadapting receptor only was mounted in a device similar to that described by Eyzaguirre and Kuffler.[35] The companion receptor, which is a thicker-fibered and adapting one, was usually cut off. The isolated preparation was suspended in a plastic box permitting dark-field illumination and was so positioned that the soma, dendrites, and muscle bundle were in a saline solution (Van Harreveld[39]), with the sensory axon rising into a layer of mineral oil, where recording electrodes picked up the discharge of impulses. Polarizing currents were applied through Ag-AgCl electrodes in contact with the preparation in different positions. In one series of experiments a

device was used which consisted of bar polarizing electrodes maintained at a constant separation and so mounted as to permit rotation with respect to the preparation (suspended from the paired forceps and recording electrodes). These bar electrodes were 1.5 cm. long and 2.5 cm. apart and provided in the central region occupied by the preparation a uniform field in one plane.

The voltage gradient was measured by micropipettes filled with 3 *M* KCl placed close to the cell at opposite sides in the plane of polarization. The cell is typically 40–60 μ long in the axono-dendritic axis, and the electrodes were typically not closer than 10 μ to the cell, no attempt being made to place them exactly in contact with the cell surface. Their tips could be seen, and the distance between them measured with a micrometer. A push-pull cathode follower input was used in conjunction with a Grass P-6/DC amplifier. This, together with a Dumont Type 322 cathode-ray oscilloscope, provided maximal amplification of 1 millivolt per inch. The nerve impulses led from the axon were monitored through a separate condenser-coupled amplifier. A galvanometer in series with the polarizing circuit measured the intensity of the imposed current, supplied by a battery and regulated with a potentiometer.

A second preparation used was the cardiac ganglion of the lobster (*Panulirus interruptus*). The ganglion was isolated completely, except for the posterior attachments to the myocardium, by which it remained suspended across a hole dissected in the myocardium. The preparation was immersed in mineral oil. The polarizing electrodes were applied to the anterior and posterior ends of the ganglion, and a pair of silver wires placed in the middle of the ganglion, separated longitudinally in the axis of polarization, provided both monitoring of the spontaneous rhythmical electrical discharge of the ganglion cells and recording of the voltage drop during the imposed polarization. The distance between these electrodes was measured with the ocular micrometer. This ganglion contains only five large neurons and four small ones (Alexandrowicz,[40] Maynard,[41, 42] Hagiwara and Bullock[43, 44]). In this case we could not place the electrodes with reference to a certain ganglion cell whose activity was identified, and we could not be sure that our estimate of threshold was the most sensitive or that the effect seen was not exerted over some small distance in the neuropile.

RESULTS

Comparative Effects of Imposed Current on Active and Inactive Neurons.—Both slow- and fast-adapting receptors, isolated from the same segment of the crayfish, were used in these experiments, and the preparations were completely immersed in mineral oil. A continuous stretch was applied to the muscle fibers, such that the nonadapting receptor was rhythmically firing between 5 and 15 impulses per second, while the adapting receptor was silent but was easily made to fire by a small increase in stretch. Polarizing electrodes were applied, one on the muscle bundles and the second on the afferent nerve close to the cell bodies. In this condition 3.6×10^{-8} ampere was sufficient to change the frequency of firing of the nonadapting receptor by 5–10 per cent. A current of 6–6.5×10^{-8} ampere changed the frequency about 35 per cent, and 1–1.2×10^{-7} ampere sufficed to alter the frequency 65–72 per cent. None of these currents induced the adapting receptor to discharge. Increasing the current to more than 20 times the threshold value for

acceleration of the already active neuron was still not sufficient to excite the inactive neuron. But still stronger currents did produce transient bursts therein. The directions of polarization to produce inhibition and acceleration were respectively cathode and anode on the nerve.

Voltage Gradient across the Soma Adequate To Modulate Frequency of an Already Active Neuron.—Using the nonadapting receptor immersed in saline, the voltage gradients measured between two microelectrodes placed close to the origin of the axon and close to the muscle bundle, respectively, in the axis of polarization, were in the most favorable cases 0.08–0.12 millivolt for effects between 5 and 25 per cent; 0.15–0.2 millivolt for effects of about 30 per cent; 0.2–0.4 millivolt for effects of more than 100 per cent (in the case of inhibition, a complete inhibition followed by rebound). These cells were *ca.* 60–80 μ in this dimension. The effective polarities were the same as those just given.

Preferential Axis of Polarization.—These values of voltage gradient were all obtained in the best axis of polarization of the neuron. When the field was rotated, a significant increase of the applied current was necessary in order to reproduce the same effect as that obtained in the axono-dendritic axis. At 25° of rotation, from 0 to 30 per cent more current was necessary. At 45°, from $1^1/_4$ up to $2^1/_2$ times as much current was necessary in most cases. At 90° rotation, in all experiments, the current needed to produce the same effect was from $1^1/_4$ up to 4 times as great. In a few cases, at 45° and 90° of rotation the polarities for acceleration and inhibition were reversed compared to those effective in the best axis. Control experiments and measurements showed that rotation did not change the threshold of current intensity by altering the position of the preparation in the field between polarizing electrodes. Changing the position of the cell without rotation did not alter the threshold current, and measuring the field showed a linear curve of voltage drop against interelectrode distance until the recording electrodes approached the polarizing electrodes within 2 mm., when a steep drop occurred.

Measurements in a Population of Neurons.—In the cardiac ganglion of the lobster, currents as low as 4.3×10^{-7} ampere are able to inhibit or accelerate by 10 per cent the frequency of the rhythmic bursts corresponding to heart beat. We did not attempt in these experiments to identify the activity of a particular one of the nine ganglion cells and measure the potential drop across it. But it may be assumed that the gradient measured between any pair of points in the ganglion separated by 100 μ will be approximately the same as between another pair of points of the same total cross-sectional area, since the impedance of the tissue is very low, and we did not encounter evidence of gross inhomogeneity in resistance. In the most favorable cases, the threshold voltage drop across 100 μ was 0.08–0.2 millivolt. The large ganglion cells are of the order of 40–60 μ in diameter of the cell body. Only the one axis of polarization was tried in this case, but the anatomy of the ganglion shows a predominance of orientation of cells in the long axis of the ganglion. The polarity, in agreement with Maynard,[45] was anterior anodale for inhibition and vice versa.

DISCUSSION

The results indicate that the output of neurons, provided that they are firing autorhythmically or presumably also if they are firing at a steady rate as a conse-

quence of a constant input, can be easily modulated in frequency by very low voltage gradients (*ca.* 0.1 mv/100 μ) measured in the external field across the whole soma, as compared to the voltage gradient necessary to stimulate a silent neuron. It will be clear in the evaluation of the values given that since they are measured across the maximum dimension, they are the maximum possible values across an effective dimension in the most effective axis of polarization. The voltage drop actually responsible for the modulation of firing frequencies can only be lower. This would be true particularly in the case in which a strategic portion of the membrane is of critical importance for the initiation of the discharge, since the value of the voltage gradient acting would be lower than that here measured in proportion to the smaller dimensions over which it acts, the total field in which the neuron lies being a relatively good conductor compared to the cell with its high-resistance membrane.

Aside from this consideration, it is obvious that even if the voltage gradient is as great as 0.1 mv/100 μ, the transmembrane potential change, resulting from the application of the polarizing current, will be smaller, something less than half of this in the maximum case. Even that value can obtain only at or near the poles of the cell, and at these two poles the effect on the resting membrane potential will be of opposite sign for a given direction of polarizing current. In the equatorial region of the cell, with respect to the axis of polarization, there will be no change in the membrane potential.

It is difficult to understand, on contemporary conventional theories of the determination of neuronal firing, how such a change in the membrane potential can influence the frequency of firing. Not only is it small, but there is in fact no change in the membrane potential of the neuron describable by a single term. Only by assuming an exceedingly critical firing level in voltage and, additionally, a limited portion of the neuron near one of the poles as being the region whose membrane potential is significant in determining firing, can the effect of the applied current be interpreted as acting through subtraction or addition from the slowly rising depolarization leading to each impulse. Of course, the critical polar region of the cell need not be a point but can be an extensive area of membrane, since the equator under consideration is not a geometrical but an electrical one and may be located at an anatomically quite polar latitude, depending on the area and the specific resistance of the membrane on each side of it. The difficulty with this hypothesis is that it requires that the very low-resistance axoplasm be at significantly different potentials in different parts of the interior of the neuron, a possibility which is certainly not excluded but for which we do not have independent evidence. Furthermore, the magnitude of the changes in steady depolarization, commonly associated with the control of firing, are of quite a different order of magnitude.[46-48] The same is true of the values of both synaptic inhibitory and excitatory potentials, which are several millivolts in amplitude, while the potentials here considered must be not greater than some tens of microvolts.

An alternative hypothesis may be that the imposed currents are acting by creating or increasing an already present gradient between different portions of the cell membrane, i.e., that the critical difference is not the change in the transmembrane potential but the change in the potential between one point on the surface and another. These possibilities are not strictly alternative or mutually exclusive, and

they suggest two new or little-considered parameters of neuronal state critically significant in determining the activity of poised or already active cells over and above other conditions which alone may be quite adequate to fire the neuron, e.g., specific chemicals or strong electric currents.

The great sensitivity of neurons to small voltage differences tends to support the views already suggested to the effect that electric field actions can play a role in the determination of probability of firing of units in ganglionic masses within the central nervous system,[49-56] at least in synchronizing cells,[6, 57-65] and that currents adequate to exert an effect are physiologically available.[7, 16, 19, 65-74]

If this is true, the considerations adduced here emphasize the significance of the morphology of the cell, since the geometric relations between the portions of the cell or dendritic membrane specially sensitive to the transmembrane changes, or the regions sensitive to difference in potential between different points on the surface with relation to the external field, acquire an overriding importance. Even more, the view here expressed emphasizes the importance of the architectural arrangements of groups of neurons and of their complex dendritic ramifications, which will be, as it were, antennae oriented with respect to the direction of currents of physiological significance to these cell masses. If this view is correct, it would be necessary to predict that in structures such as the cerebral cortex, in which orientation of many neurons is not random but characteristic, these neurons must normally be subject to physiologically significant field effects of a relatively constant axis. Since this axis is probably the most effective axis for polarization effects and since we may guess, for example in the case of the pyramidal cells of the cerebral cortex, that this will be the longitudinal axis of the cell, the expectation stated means that we would anticipate a higher voltage gradient in the extracellular fluid consequent to neuronally significant events in the axis normal to the surface than tangentially. But different axes may be simultaneously effective, as is suggested by the specific geometry of the processes in the molecular layer of the cerebellar cortex.

Finally, it may be pointed out that this view, while superficially in agreement with the theory of Gesell,[75, 76] is significantly different in this respect: that the gradient between one part of the surface and another, which is supposed to be important in determining neuronal firing, exists only in the already active cell or, in the active cell, is sensitive to changes of a different order of magnitude. The silent cell, even when the attempt is made to poise it close to its threshold, as in the adapting stretch receptor, requires enormously more voltage gradient to be fired than the active cell does to be modulated.

SUMMARY

1. The voltage in the extracellular field across a single nerve cell has been measured with microelectrodes during polarization with imposed current just sufficient to modify the frequency of firing of an already active cell, using a preparation from the nonadapting stretch receptor of the crayfish abdomen and another from the cardiac ganglion of the lobster.

2. It is argued that, unless the current density can be determined across the strategic portion of the cell membrane, this is the most suitable measure of the intensity of imposed currents which can exert an effect on the probability of firing of a neuron. The transmembrane potential does not provide such a measure.

3. In the most effective axis of polarization, it was found that a voltage gradient in the neighborhood of 0.1 mv/100 μ markedly influenced active cells. Currents of more than 20 times this value are required to fire a silent cell, even if it has been poised, i.e., the adapting stretch receptor, under a physiological degree of stretch.

4. The actual value of the critical voltage drop across the essential structure can only be smaller than this, and it may be much smaller. The transmembrane potential change with polarization may be from 0 to nearly one-half this value and is thus, at the most, several orders of magnitude smaller than the changes in membrane potential for threshold electrical stimulation or for several cases cited from the literature associated with alteration in activity or with inhibition or excitation.

5. The findings may mean that exceedingly critical firing levels in voltage across the membrane exist which would have to be confined to a certain portion of the cell near one of the poles relative to the effective axis of polarization. Alternatively, it is supposed that the imposed current acts by creating or increasing a gradient between different portions of the cell membrane, not across it but between one point on the surface and another. These possibilities are not mutually exclusive, and they suggest two new or little-considered parameters of neuronal state critically significant in determining activity over and above other conditions.

6. The great sensitivity of neurons to small voltage differences supports the view that electric field actions can play a role in the determination of probability of firing of units in ganglionic masses, in response to physiologically available currents. If true, this conclusion emphasizes the significance of morphology of the cell and of architectural arrangement of groups of neurons and their dendritic ramifications.

* Aided by grants from the National Institute of Neurological Diseases and Blindness (B21) and the University of California.

[1] G. Fritsch and E. Hitzig, *Arch. Anat. Physiol. Med.*, **37**, 300–322, 1870.
[2] D. H. Barron and B. H. C. Matthews, *J. Physiol.*, **92**, 276–321, 1938.
[3] H. Dubner and R. W. Gerard, *J. Neurophysiol.*, **2**, 142–152, 1939.
[4] F. Scheminzky, *Pflügers Arch. ges. Physiol.*, **243**, 439–456, 1940.
[5] F. Bremer, *Arch. intern. Physiol.*, **51**, 51–84, 1941.
[6] F. Bremer, *ibid.*, pp. 211–260, 1941.
[7] B. Libet and R. W. Gerard, *J. Neurophysiol.*, **4**, 438–455, 1941.
[8] F. Bremer and V. Bonnet, *Arch. intern. Physiol.*, **52**, 153–194, 1942.
[9] F. Bremer, V. Bonnet, and J. Moldaver, *Arch. intern. Physiol.*, **52**, 215–248, 1942.
[10] A. Van Harreveld, M. S. Plesset, and C. A. G. Wiersma, *Am. J. Physiol.*, **137**, 39–46, 1952.
[11] T. H. Bullock, H. S. Burr, and L. F. Nims, *J. Neurophysiol.*, **6**, 85–98, 1943.
[12] R. Granit, *J. Physiol.*, **105**, 45–53, 1946.
[13] F. Scheminzky, *Experientia*, **3**, 169–176, 1947.
[14] C. G. Skoglund, *Acta physiol. Scand.*, **14**, Suppl. 47, 1–13, 1947.
[15] G. H. Bishop and T. L. O'Leary, *EEG clin. Neurophysiol.*, **2**, 401–406, 1950.
[16] A. Mollica and H. Terzian, *Boll. Soc. ital. biol. sper.*, **26**, 1244–1245, 1950.
[17] C. Ajmone-Marsan, M. G. F. Fuortes, and F. Marossero, *J. Physiol.*, **113**, 316–321, 1951.
[18] S. Goldring and J. L. O'Leary, *J. Neurophysiol.*, **14**, 275–288, 1951.
[19] T. Otani, S. Iwata, and T. Furukawa, *Japan. J. Physiol.*, **2**, 154–165, 1951.
[20] J. M. Brookhart and P. H. Blanchly, *Am. J. Physiol.*, **171**, 711, 1952.
[21] H. K. Hartline, N. A. Culter, Jr., and H. G. Wagner, *Federation Proc.*, **11**, 65–66, 1952.
[22] I. Tasaki and C. Fernandez, *J. Neurophysiol.*, **15**, 497–512, 1952.
[23] J. Alanis, *J. Physiol.*, **120**, 569–578, 1953.
[24] O. Lowenstein, *Nature*, **71**, 549–550, 1953.

[25] A. Mollica, G. Moruzzi, and R. Naquet, *EEG clin. Neurophysiol.*, **5**, 571–584, 1953.
[26] A. Mollica and G. F. Rossi, *Arch. Fisiol.*, **54**, 219–230, 1954.
[27] M. G. F. Fuortes, *J. Physiol.*, **126**, 494–506, 1954.
[28] C. A. Terzuolo, *Arch. intern. Physiol.*, **62**, 179–196, 1954.
[29] C. Edwards, *J. Physiol.*, **127**, 636–640, 1955.
[30] V. Bonnet, *Arch. intern. Physiol. Biochem.*, **64**, 141–167, 1956.
[31] R. V. De Vito, A. Brusa, and A. Arduini, *J. Neurophysiol.*, **19**, 241–253, 1956.
[32] J. S. Alexandrowicz, *Quart. J. Microscop. Sci.*, **92**, 163–199, 1951.
[33] C. A. G. Wiersma, E. Furshpan, and E. Florey, *J. Exptl. Biol.*, **30**, 136–150, 1953.
[34] S. W. Kuffler, *J. Neurophysiol.*, **17**, 558–574, 1954.
[35] C. Eyzaguirre and S. W. Kuffler, *J. Gen. Physiol.*, **39**, 87–119, 1955.
[36] C. Eyzaguirre and S. W. Kuffler, *ibid.*, pp. 121–153, 1955.
[37] S. W. Kuffler and C. Eyzaguirre, *J. Gen. Physiol.*, **39**, 155–184, 1955.
[38] E. Florey and E. Florey, *J. Gen. Physiol.*, **39**, 69–85, 1955.
[39] A. Van Harreveld, *Proc. Soc. Exptl. Biol. Med.*, **34**, 428–432, 1936.
[40] J. S. Alexandrowicz, *Quart. J. Microscop. Sci.*, **75**, 182–250, 1932.
[41] D. M. Maynard, Jr., *Biol. Bull.*, **104**, 156–170, 1953.
[42] D. M. Maynard, Jr., *ibid.*, **109**, 420–436, 1955.
[43] S. Hagiwara and T. H. Bullock, *Biol. Bull.*, **109**, 341, 1955.
[44] S. Hagiwara and T. H. Bullock, *J. Gen. Physiol.*, 1956 (in press).
[45] D. M. Maynard, Jr., Ph.D. thesis, University of California at Los Angeles, 1954.
[46] T. Araki and T. Otani, *J. Neurophysiol.*, **18**, 472–485, 1955.
[47] H. K. Hartline, H. G. Wagner, and E. F. MacNichol, Jr., *Cold Spring Harbor Symposia Quant. Biol.*, **17**, 125–141, 1952.
[48] B. Katz, *J. Physiol.*, **111**, 216–286, 1950.
[49] E. D. Adrian and B. H. C. Matthews, *J. Physiol.*, **81**, 440-471, 1934.
[50] E. D. Adrian, *J. Physiol.*, **88**, 127–161, 1936.
[51] R. W. Gerard, *Cold Spring Harbor Symposia Quant. Biol.*, **4**, 292–304, 1936.
[52] G. H. Bishop, *Ann. Rev. Physiol.*, **3**, 1–20, 1941.
[53] R. W. Gerard, *Ohio J. Sci.*, **41**, 160–172, 1941.
[54] R. W. Gerard, *Schweiz. med. Wochschr.*, **71**, 397–398, 1941.
[55] A. Fessard, "Mechanisms of Nervous Integration and Conscious Experience," in E. D. Adrian, F. Bremer and H. H. Tasper, *Brain Mechanisms and Consciousness* (Springfield, Ill.: Charles C Thomas, 1954).
[56] T. H. Bullock, "Neuronal Integrative Mechanisms," in *Recent Advances in Invertebrate Physiology*, ed. B. T. Schneer (Eugene: University of Oregon Press, 1956).
[57] E. C. Adrian and S. Matthews, *J. Physiol.*, **65**, 273–298, 1928.
[58] E. C. Adrian, *The Mechanism of Nervous Action* (Philadelphia: University of Pennsylvania Press, 1931).
[59] R. W. Gerard and B. Libet, *Am. J. Psychiat.*, **19**, 1125–1152, 1939.
[60] R. W. Gerard and B. Libet, *Melanges Ozorio de Almeida, Rio de Janeiro*, pp. 288–294, 1939.
[61] A. Arvanitaki, *Arch. intern. Physiol.*, **52**, 381–407, 1942.
[62] A. Fessard, *Compt. rend. soc. biol.*, **136**, 114–117, 1942.
[63] A. Fessard, *ibid.*, pp. 268–270, 1942.
[64] T. Gualtierotti, *J. Physiol.*, **121**, 106–116, 1953.
[65] C. A. Terzuolo and B. E. Gernandt, *Am. J. Physiol.*, 1956 (in press).
[66] T. H. Bullock, *Yale J. Biol. and Med.*, **7**, 657–679, 1945.
[67] S. Goldring and J. L. O'Leary, *EEG clin. Neurophysiol.*, **3**, 329–340, 1951.
[68] T. Gualtierotti, *J. Physiol.*, **118**, 361–372, 1952.
[69] Chou-Loh Li and H. Jasper, *J. Physiol.*, **121**, 117–140, 1953.
[70] S. Goldring and J. L. O'Leary, *EEG clin. Neurophysiol.*, **6**, 189–200, 1954.
[71] S. Goldring and J. L. O'Leary, *ibid.*, pp. 201–212, 1954.
[72] M. Dondey and R. S. Snider, *EEG clin. Neurophysiol.*, **7**, 265–272, 1955.
[73] E. F. Vastola, *EEG clin. Neurophysiol.*, **7**, 557–567, 1955.
[74] W. G. Walter, *EEG clin. Neurophysiol.*, **8**, 162, 1956.
[75] R. Gesell, *Ergeb. Physiol.*, **43**, 477–639, 1940.
[76] R. Gesell, *Schweiz. med. Wochschr.*, **71**, 398–400, 1941.

*IN VITRO AND IN VIVO EFFECTS OF A NERVE GROWTH-STIMULATING AGENT ISOLATED FROM SNAKE VENOM**

By Rita Levi-Montalcini and Stanley Cohen

DEPARTMENT OF ZOÖLOGY, WASHINGTON UNIVERSITY, ST. LOUIS, MISSOURI

Communicated by V. Hamburger, July 19, 1956

INTRODUCTION

Previous reports have dealt with a nerve growth-stimulating effect of mouse sarcomas 180 and 37 on sensory[1] and sympathetic ganglia of the chick embryo.[2] The effect consisted in a striking increase in size and cell number of sensory and sympathetic ganglia adjacent to the tumor and of an extensive invasion of their fibers into the tumor. Transplantation of the tumor onto the allantoic membrane had a growth-promoting effect similar to the intra-embryonic transplantations.[3, 4] In the allantoic transplants the tumor was not accessible to nerve fibers; the effect was therefore mediated through the circulation which the tumor shared with its host. Embryos bearing mouse tumors in the abdominal wall or on the allantoic membrane differed from controls also in another respect: the viscera were overloaded with sympathetic nerve fibers, whereas control embryos showed practically no visceral innervation at corresponding stages of development. Large nerve bundles were also found in blood vessels, where they formed agglomerates projecting into the lumen of the vessel. The agent released by sarcomas 180 and 37 has therefore a dual effect: it stimulates the overgrowth of sensory and sympathetic ganglia and is responsible for the aberrant distribution of sympathetic nerves in the viscera and blood vessels. Two alternative hypotheses were offered: either the agent affects primarily the growth and differentiation processes of nerve cells, or changes of the periphery are the primary effect and the overgrowth of the ganglia is an indirect effect.[4] The following experimental result favored the first hypothesis.

Sensory and sympathetic ganglia were explanted in vitro in proximity to a fragment of mouse tumor. Under these conditions the ganglia are directly exposed to the tumor, and influences of the organism are excluded. Ganglia explanted together with normal embryonic tissues of chick or mouse served as controls. Whereas very few nerve fibers emerged from the control ganglia in the first 24 hours, a dense fibrillar "halo" surrounded the ganglia exposed to the tumor at the end of the first day.[5] These experiments indicated a direct action of the neoplastic agent on nerve cells. The tissue-culture method was then used to test the effect of a cell-free homogenate of the tumor. When it was found to possess the same activity as the growing tumor in vitro, the investigation was directed toward the purification and identification of the agent.[6, 7]

In the course of this work another growth-stimulating agent was discovered in snake venoms. The properties of this agent will be dealt with in another communication.[8] It is the object of the present report to describe its biological effects on sensory and sympathetic ganglia in vitro and in vivo and to compare them with those of the tumor factor.

EXPERIMENTS IN VITRO

The investigation was centered on sensory ganglia of 7-day chick embryos. The standard hanging-drop technique was used. The medium consisted of one part plasma, one part synthetic medium, and one part of the snake venom to be assayed. We used crude venom from *Crotalus adamanteus* (rattlesnake) and *Agkistrodon piscivorus* (moccasin) and a protein fraction (DEAE fraction) derived from the moccasin venom.[8] In each culture three to four ganglia were explanted. A total of several hundred ganglia were tested. A series of controls was run in which one part of an isotonic saline solution was added to the medium instead of the snake-venom fraction. The cultures were incubated at 37° and examined after 6, 12, and 24 hours.

Results.—The addition to the medium of 7 μg/ml of the crude snake venom or of 1 μg/ml of the DEAE fraction had identical effects. The following description therefore applies to the experiments performed with both. The results were in fact remarkably consistent, and only minor fluctuations were observed.

Six hours after the beginning of the experiment, a dense "halo" of short nerve fibers surrounded the ganglia. The fibers increased rapidly in length and density in the following hours. At 24 hours the fibrillar halo had reached its maximal density (Fig. 1). The characteristics of this most unusual fiber outgrowth were identical with those described in previous investigations dealing with mouse tumor;[5, 6] they will therefore not be recounted here (Figs. 3 and 4). It may only be mentioned that the cultures differed from controls also in another respect. Whereas in control cultures the fiber outgrowth was very scanty in the first 24 hours and the nerve fibers grew in a wavy, irregular fashion (Fig. 2), in the present experiments the fibers grew straight, radially, and showed extensive branching, forming a brushlike border at their tips. In most cases, a second, shorter halo formed in the clot between the twelfth and the twenty-fourth hour. As in the previous experiments with tumor, very few or no spindle cells were found among the nerve fibers in the first 24 hours. In the following days, however, a large number of these cells migrated from the explant between the fibers. In the present experiments a considerable degree of liquefaction of the plasma occurred. As a result, a disintegration of the fibrillar halo was observed in many of the cultures after 48–72 hours. The cultures were then discontinued.

A more limited number of experiments were performed with sympathetic ganglia explanted from 8- to 9-day chick embryos. The results were in all respect similar to the ones described above.

EXPERIMENTS IN VIVO

It had been suggested previously that the in vitro and in vivo effects of tumors were produced by the same agent. However, preliminary experiments of injections of active tumor fractions into the yolk of 7-day embryos were negative. We attribute this failure to the fact that the pure fraction was available only in very small amounts and that probably the quantities injected were considerably smaller than those produced continually by living tumor. The striking similarity between the in vitro effects of the tumor and of the snake venom, and the fact that the activity of the purest fraction of snake venom is 1,000 times that of the purest tumor fraction (on a dry-weight basis), suggested a repetition of the injection experiments with snake venom.

The effect of the venom was assayed by injecting it into the yolk of embryos ranging between 6 and 8 days of development. A small window was opened in the shell, and the fraction DEAE of the venom (containing 150 μg protein/ml together with 100 μg/ml of terramycin) was injected into the yolk with a 1-ml. syringe.

In these preliminary experiments the total amount of venom injected, the stage of the embryos, and the length of time of the experiments were varied, in order to determine (*a*) the tolerance of the embryo, (*b*) the stages of development which are sensitive to the effect of the venom, and (*c*) the length of time required to secure an effect.

The injections were repeated two or three times a day. A total of 120 embryos were injected; 17 were recovered and impregnated with silver following the Cajal–De Castro technique. They were sectioned serially and used for the present investigation.

Results.—The toxicity of the venom was reflected in the severe mortality and in the diffuse edema which affected the surviving embryos. The results proved beyond doubt that the venom possesses a growth-stimulating effect on sensory and sympathetic ganglia of the embryo similar in every respect to the effect of mouse sarcomas. The increase in size of the sympathetic system and, in particular, of its prevertebral chain was so striking that it overshadowed the reaction of the sensory ganglia. The same results were obtained in embryos bearing mouse sarcomas. The analysis was then focused on the sympathetic system.

Of the 17 embryos, 11 had been injected over a period of between 36 hours and 4 days, and 6 over a period of only 26 hours. Ten of the 11 embryos injected for a longer period showed a considerable increase in size of the sympathetic ganglia and an atypical distribution of nerve fibers in the viscera. The intensity of the effect varied in different embryos: it increased parallel to the length of the time the embryos had been exposed to the venom. A definite increase in the prevertebral sympathetic complex was observed in one embryo injected over a period of 36 hours, from 6 to $7^1/_2$ days, with a total amount of 0.5 ml. Nerve bundles emerging from the prevertebral complex entered the adjacent mesonephroi (Figs. 9 and 10). No nerves are found in the mesonephroi at corresponding stages in control embryos. The nine other embryos which showed a positive effect were fixed 2, 3, and 4 days after the beginning of the experiment. The total amounts of venom injected were: 0.6 ml. in one, 1.3 ml. in four, 1.8 ml. in two, and 3.0 ml. in two cases. A maximal effect was observed in the two embryos injected over a period of 4 days with 3.0 ml. and recovered at 10 days. The embryos injected for only 26 hours showed no effect.

In one of the embryos injected over a period of 4 days, one ganglion of the prevertebral sympathetic complex at the level of the adrenal was measured. Each section was projected with the help of the camera lucida, and the contour of the ganglion was drawn on cardboard. The total number of sections was then weighed, and the weight was compared to that of the same prevertebral ganglion of a control embryo.

The results gave an increase in weight of 3.2 times the control. Since the weight is proportional to the volume, the figure indicates a similar increase in volume. Estimates of cell numbers were obtained by inserting a micrometer disk in the ocular and counting the number of cells in 200 squares each in comparable regions of ganglia from normal and injected embryos, using oil immersion. The results are

1,066 cells in 200 squares of the injected embryo and 2,158 in the control. The ratio of cell number per unit area was therefore 1:2. The volume ratios were then divided by the cell-count ratios to obtain an estimate of the relative increase in cell numbers. The latter amounts to 1.6. Hence the over-all increase in size is due to a greater extent to cellular hypertrophy and to a lesser extent to an increase in cell numbers. Comparable results were obtained in embryos bearing tumors.[2]

A comparison of hyperplastic and normal ganglia (Figs. 5 and 6) shows that the cells of the injected embryos are not only larger than the controls but also in a more advanced stage of differentiation. This is attested by the presence of neurofibrils in the cytoplasm of almost all the sympathetic neurons in the experimental cases, but only in very few of the controls (Figs. 7 and 8). Furthermore, the nerve bundles which emerge from the hyperplastic and hypertrophic ganglia are much larger than those in the controls.

COMPARISON OF THE IN VIVO EFFECTS OF MOUSE SARCOMAS AND SNAKE VENOM

The results of the preliminary experiments with snake venoms reported here point to a striking similarity between the two phenomena.

In both groups of experiments not only the mesonephroi (Figs. 9 and 10) but also other viscera—thyroid, metanephroi, spleen, and liver—are invaded by sympathetic fibers. In the experiments with snake venom as in those with tumor, nerve bundles perforated the wall of thoracic veins and projected inside the blood vessels (Figs. 11 and 12). Since the experiments with venom were not performed in as late developmental stages as those with tumor, the neuromas were of small size and comparable only to neuromas found in 9–10-day embryos bearing a mouse tumor. The quantitative data reported above are also in line with those given for tumor effects.[2, 4]

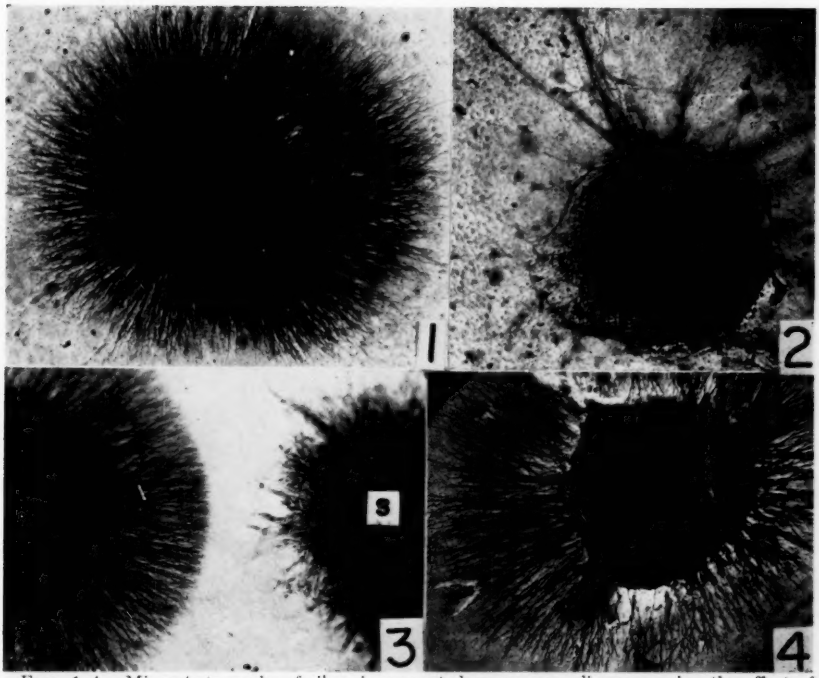
In both groups of experiments no morphological changes were detected in any part of the nervous system other than spinal and sympathetic ganglia.

CONCLUDING REMARKS AND SUMMARY

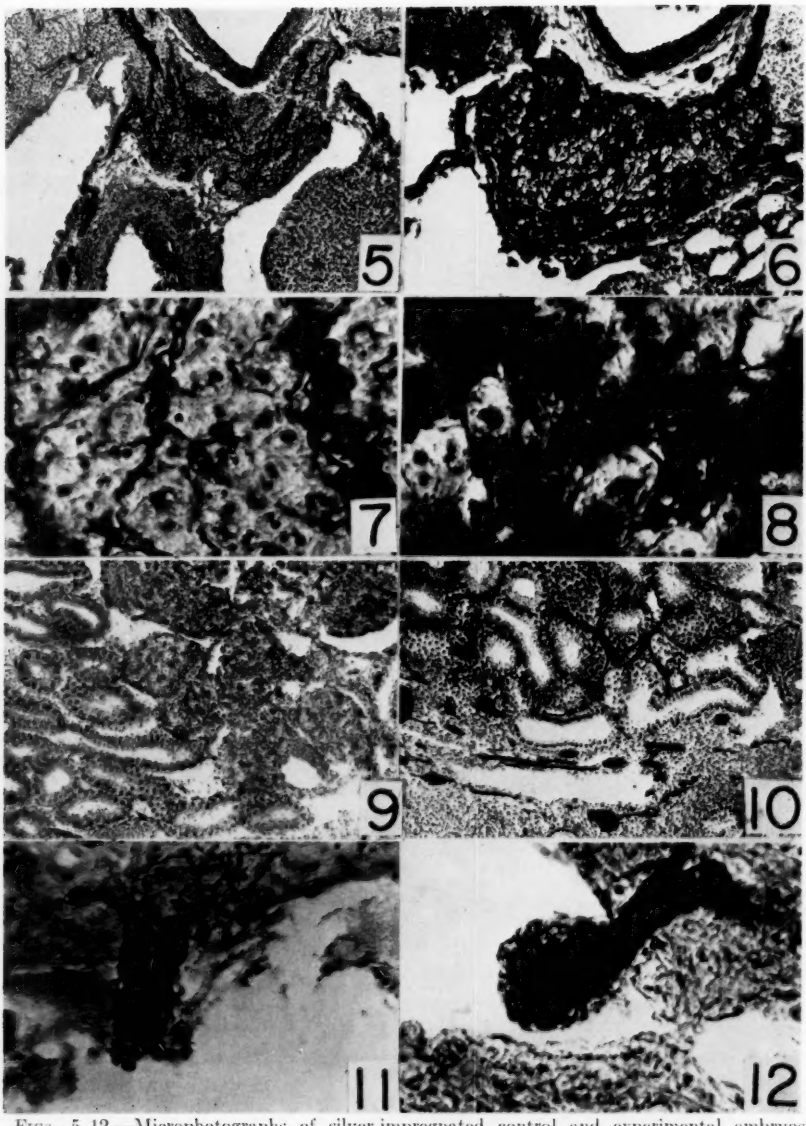
A nerve growth-promoting agent was discovered in snake venoms, and its in vitro and n vivo effects were investigated.

In both sets of experiments the nerve fiber outgrowth from sensory and sympathetic ganglia was found to be far beyond the normal range. Furthermore, the experiments in vivo gave evidence of a remarkable increase in cell size and cell numbers in sympathetic ganglia and of an acceleration of their differentiative processes. The nerve fibers produced in excess entered into the adjacent viscera.

These observations show that the snake venoms duplicate the effects of a "protein fraction" of mouse sarcomas in vitro and the effect of living sarcomas in vivo. The similarity of the two agents is further emphasized by the data which were obtained in attempts to purify them. In both instances, we are apparently dealing with a protein or a substance bound to protein.[7, 8] The purest fraction of snake venom obtained so far is approximately 1,000 times as effective as our purest fraction of mouse tumor (on a dry-weight basis).[8] This fact as well as the ready availability of different snake venoms will facilitate the further analysis of the phenomena reported above.



FIGS. 1–4.—Microphotographs of silver-impregnated sensory ganglia, comparing the effect of the snake venom, of the intact tumor, of the cell-free extract of the same tumor, and a control ganglion. All ganglia were isolated from 7-day embryos and cultured for 24 hours. Fig. 1, sensory ganglion growing in a medium to which the fraction DEAE of moccasin venom had been added. Fig. 2, control ganglion. Fig. 3, ganglion combined with a fragment of sarcoma 37 (*S*). Fig. 4, ganglion growing in a medium to which cell-free extract of the tumor has been added.



FIGS. 5–12.—Microphotographs of silver-impregnated control and experimental embryos. Fig. 5, prevertebral sympathetic ganglion of a 10-day embryo. Fig. 6, prevertebral sympathetic ganglion of a 10-day embryo injected with moccasin snake venom (fraction DEAE) for 4 days. Figs. 7 and 8, enlarged sectors of Figs. 5 and 6, respectively. Figs. 9 and 10, mesonephros of a control and of an injected 10-day embryo. Notice the sympathetic nerve fibers between the mesonephric tubules in the injected embryo. Fig. 11, neuroma projecting into the lumen of a vein of a 9-day embryo with an extra-embryonic sarcoma 37. Fig. 12, same, in a 10-day embryo, injected with moccasin venom, fraction DEAE (see text).

Our failure, so far, to obtain positive effects after in vivo injection of mouse tumor fractions was attributed to the relative weakness of the tumor fraction rather than to the presence of two different agents in the tumor—one that would be responsible for the in vivo effects and one responsible for the in vitro effects. The present finding that the much stronger venom factor reproduces both effects in every detail gives further strong support to our contention.

* This work has been supported by Grant B-463 from the National Institute of Neurological Diseases and Blindness of the National Institutes of Health, Public Health Service, by a grant of the American Cancer Society, recommended by the Committee on Growth, National Research Council, and by a contribution from an institutional grant of the American Cancer Society to Washington University. The able assistance of Miss Barbara Bankhead is gratefully acknowledged.

[1] E. D. Bueker, *Anat. Record*, **102**, 369–390, 1948.
[2] R. Levi-Montalcini and V. Hamburger, *J. Exptl. Zool.*, **116**, 321–362, 1951.
[3] R. Levi-Montalcini, *Ann. N.Y. Acad. Sci.*, **55**, 330–343, 1952.
[4] R. Levi-Montalcini and V. Hamburger, *J. Exptl. Zool.*, **123**, 233–288, 1953.
[5] R. Levi-Montalcini, H. Meyer, and V. Hamburger, *Cancer Research*, **14**, 49–57, 1954.
[6] S. Cohen, R. Levi-Montalcini, and V. Hamburger, these PROCEEDINGS, **40**, 1014–1018, 1954.
[7] S. Cohen and R. Levi-Montalcini, *Cancer Research* (in press).
[8] S. Cohen and R. Levi-Montalcini, these PROCEEDINGS, **42**, 571–574, 1956.

ON TIME LAGS IN EQUATIONS OF GROWTH

BY P. J. WANGERSKY AND W. J. CUNNINGHAM

OSBORN ZOÖLOGICAL LABORATORY; DUNHAM LABORATORY OF ELECTRICAL ENGINEERING, YALE UNIVERSITY, NEW HAVEN, CONNECTICUT

Communicated by G. E. Hutchinson, July 27, 1956

An earlier paper by one of the authors[1] analyzed the equation

$$\frac{dN_{(t)}}{dt} = bN_{(t)}\left[\frac{K - N_{(t-\tau)}}{K}\right], \tag{1}$$

which has been suggested[2] as a mathematical description of a population growing at a constant reproductive rate (b) toward a saturation value (K), with a self-regulatory system represented by $(K - N)/K$ which takes a finite time (τ) to react to changes in the environment caused by the approach of the population to the saturation level.

When this equation is used to describe real populations, a second time lag, although not expressed, is clearly implied. In the species of animals normally used for laboratory experiments in population dynamics a finite amount of time is needed for the animal to respond to changes in its environment. The lag introduced into the self-regulatory system in equation (1) becomes of great importance as (N) approaches (K). While the population is still in the exponential stage of its growth, and (N) is quite small in comparison to (K), another lag, the time needed for the reproductive process to take place, is more important.

The equation as written states that the change in the number of animals in the time interval (dt) is some function of the number of animals present at time (t).

This function is a statistical estimate of the proportion of animals present at time (t) which were beginning the process of reproduction at some previous time such that they will give birth at time $(t + dt)$. As long as the population is in the exponential part of the growth curve, the time involved in the process of reproduction is sufficiently constant in a given species that it may be expressed as a constant lag, (τ_1). Actually, even in this part of the growth curve there is considerable variation between individuals in reproductive time lag. The results of this variation will be discussed later. The equation of growth can be written with this lag expressed, rather than implied, as

$$\frac{dN_{(t)}}{dt} = bN_{(t-\tau_1)}\left[\frac{K - N_{(t-\tau_2)}}{K}\right], \tag{2}$$

or, when simplified by combining constants, as

$$\frac{dN_{(t)}}{dt} = bN_{(t-\tau_1)} - cN_{(t-\tau_1)}\,N_{(t-\tau_2)}, \tag{3}$$

where $c = b/K$ and τ_2 = reaction time lag.

For this equation there are two equilibrium, or singular, values of N at which $dN_{(t)}/dt = 0$. These values are (1) $N = N_1 = 0$ and (2) $N = N_2 = b/c = K$. In studying the solutions for equation (3), it is most convenient first to normalize the equation, using the definitions $p = t/\tau_1$, $q = \tau_2/\tau_1$, and $r = N/N_2$. In terms of these quantities, equation (3) becomes

$$r'_{(p)} = b\tau_1[r_{(p-1)} - r_{(p-1)}r_{(p-q)}], \tag{4}$$

where primes indicate differentiation with respect to (p). At singularity (2), (r) has the value unity. Near this singularity the solution may be studied by replacing (r) with $(1 + u)$, where (u) is a small quantity. The resulting variational equation is

$$u'_{(p)} = b\tau_1[-u_{(p-q)}], \tag{5}$$

where a product term is neglected. If (q) is small, it is a fair approximation to write

$$u_{(p-q)} = u_{(p)} - qu'_{(p)} + {}^1/_2q^2u''_{(p)}, \tag{6}$$

which, combined with equation (5), gives

$$u'' + \frac{2}{q^2}\left[\left(\frac{1}{b\tau_1} - q\right)u' + u\right] = 0. \tag{7}$$

Equation (7), approximately valid near singularity (2), indicates that, if $q = 1/b\tau_1$, or $b\tau_2 = 1$, a steady-state oscillation will occur. The period of this oscillation in real time is $T = 2\pi\tau_1/(2/q^2)^{1/2} = \sqrt{2}\pi\tau_2$. That is, if values of (b) and (τ_2) are large enough, the population will oscillate in a cyclic manner about the saturation value with a period approximately 4.4 times the reaction time lag. Furthermore, if $[(2/q^2)\,(1/b\tau_1 - q)]^2 = 4(2/q^2)$, or $b\tau_2 = \sqrt{2} - 1$, critical damping exists. Populations with slow rates of growth or very fast reaction times will approach the saturation value monotonically. Intermediate values of (b) and (τ_2) give rise to damped

oscillation about the saturation value. This would seem to be the type of curve generally observed in laboratory growth experiments. Utida[3] found such a damped oscillation in populations of the azuki bean weevil. Gause's[4] data on the growth of *Paramecium aurelia* and *P. bursaria* in separate cultures suggest this type of oscillation. Unfortunately, the cultures were not carried long enough or observed often enough to allow quantitative use of his data.

Near singularity (1), the variational equation for the solution is

$$u'_{(p)} = b\tau_1[u_{(p-1)}], \tag{8}$$

which is approximately equal to

$$u' = \left[\frac{b\tau_1}{1 + b\tau_1}\right]u. \tag{9}$$

The solution here is a growing exponential curve, with the growth rate in terms of real time being $b/(1 + b\tau_1)$.

The result of this analysis is that near $N = N_1 = 0$ the solution depends primarily upon the product $(b\tau_1)$. The presence of the delay (τ_1) makes the initial growth rate less than (b), which it would be if $(\tau_1) = 0$, by the factor $1/(1 + b\tau_1)$. Near $N = N_2 = K$, the solution depends primarily on the product $(b\tau_2)$. Thus, at the beginning of growth, while the population is increasing exponentially, the change in N depends largely on the reproductive time lag. As the population nears its saturation value, the governing factor becomes (τ_2), the reaction lag. In many cases in nature we are dealing with populations near their equilibrium levels, and we may safely disregard the reproductive lag. However, in populations such as plankton and bacteria, which are repeatedly reduced to very small numbers by exhaustion of the medium or by seasonal changes in physical factors, the reproductive lag may be more important than the reaction lag in describing the population.

Since reproduction in animals is a discontinuous process, while the growth equations are continuous, the agreement between theory and experiment at the beginning of population growth can be no better than fair. The mathematician can construct fractional animals somewhat more easily than can the experimentalist. For those species whose generations do not overlap, any continuous function can be only an approximate description and a working guide. In most species of animals, however, there is enough overlap of generations and enough individual variability in reproductive lag so that even a population started from one pair of animals will behave as a continuous function before very many generations have passed.

The effect of reproductive lag on a population can probably be demonstrated most readily in extremely small populations and in the very beginning of population growth. Many investigators have found a definite lag in the growth rate of newly started populations and have started their mathematical descriptions at the end of the lag period. This lag has sometimes been attributed to the necessity for the animal to "condition" the medium and to other external causes. In the case of these animals living in a liquid medium, this might mean the removal of traces of metal ions by chelation or adjustment of the pH by the production of metabolic waste products. Some part of the lag must be internal, due to the reproductive lag discussed above. Differences in the physiological state of the animals when

they are added to the culture could account for differences in lag in separate experiments. It should be possible to test this in the laboratory.

Oscillation about N_2 may occur if $(b\tau_2)$ is large enough. It is possible that the intrinsic oscillations in populations postulated by Slobodkin[5] are due at least in part to the time lags inherent in the physiology of the animals. When the lags are of sufficient size, such oscillations must occur but may be hidden by the effects of external events.

The approximation for $u_{(p-q)}$ used in equation (6) is sufficiently poor that numerical values of (b) and (τ_2) for critical damping and for steady-state oscillations found from equation (7) are somewhat in error, being too small. The differences here are quite similar to those found in studying equation (1) with $(\tau_1) = 0$, shown in an earlier paper.[1]

Of the three types of solutions found—the slow monotonic approach to saturation, the damped oscillation, and the limit cycle—the damped oscillation would seem to be the most advantageous to a species in an evolutionary sense. An oscillation of some sort, with continuous removal of a part of the population, would serve to keep the population structure from becoming overloaded with older members. The oscillation would serve somewhat the same purpose as the "jitter" which is built into some mechanical systems to remove the necessity for overcoming friction when changing from one level of activity to another. If some sort of oscillation is indeed an advantage to the species, a damped oscillation with its faster recovery from disturbances would be far preferable to a cyclical one. In order for a limit cycle generated in this manner to be of small amplitude, very careful balancing of (b) and (τ_2) is needed. The usual limit cycle approaches zero fairly closely and so would be particularly vulnerable to disturbances.

The authors are greatly indebted to Professor G. E. Hutchinson and Dr. Gordon A. Riley for their many comments and suggestions.

[1] W. J. Cunningham, these PROCEEDINGS, **40**, 708–713, 1954.
[2] G. E. Hutchinson, *Ann. N.Y. Acad. Sci.*, **50**, 221–246, 1948.
[3] S. Utida, *Mem. Coll. Agr., Kyoto Imp. Univ.*, **51**, 27–34, 1941.
[4] G. F. Gause, *La Théorie mathématique de la lutte pour la vie* (Paris: Hermann & Cie, 1935).
[5] L. B. Slobodkin, *Ecol. Monographs*, **24**, 69–88, 1954.

ERRATUM: COURSE OF PRODUCTION OF AN ISOANTISERUM EFFECTING TUMOR HOMOGRAFT SURVIVAL IN MICE

In the article of the above title appearing in these PROCEEDINGS, **42**, 269–273 1955, the following correction should be made:

On page 270, column 2, line 4 of Table 1, for "20/20" read "20/22".

NATHAN KALISS

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of the PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

MANUSCRIPTS will be accepted only from members of the Academy or from chairmen of the divisions of the National Research Council who will assume responsibility for the propriety and scientific standards of the paper and should be addressed to *Proceedings of the* NATIONAL ACADEMY OF SCIENCES, Attention Miss Mary D. Alexander, University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois. Articles of which a member of the Academy is author or co-author may extend to 10 printed pages, those by non-members to 3 printed pages; space above these limits will be billed to the author at a rate based on the actual composition costs of the article. Accompanying the script should be a statement by the author or his representative that he assumes responsibility for the charges if incurred.

GALLEY PROOF will be sent, with the understanding that charges for author's corrections shall be billed to him. Author should therefore make final revisions on the typewritten manuscripts. Page proofs will not be sent.

REPRINTS should be ordered when proof is returned, on the reprint order form which will be inclosed with the proof. The order blank gives a tabular estimation of the cost of reprints; a copy will be supplied at any time upon request.

CONTENTS

Page

BIOCHEMISTRY.—A NERVE GROWTH-STIMULATING FACTOR ISOLATED FROM SNAKE VENOM . *By Stanley Cohen and Rita Levi-Montalcini* 571

BIOCHEMISTRY.—BACTERIAL PROTOPLASTS INDUCED BY PENICILLIN . *By Joshua Lederberg* 574

BIOCHEMISTRY.—HINDERED *Cis* ISOMERS OF VITAMIN A AND RETINENE: THE STRUCTURE OF THE NEO-B ISOMER . . . . . *By William Oroshnik, Paul K. Brown, Ruth Hubbard, and George Wald* 578

BIOCHEMISTRY.—ON THE SYNTHESIS OF ANTIBODY PROTEIN IN VITRO . . . . . *By Donald F. Steiner and H. S. Anker* 580

BIOCHEMISTRY.—PRODUCTION OF PROTOPLASTS OF *Escherichia Coli* BY LYSOZYME TREATMENT . . . . . *By Norton D. Zinder and William F. Arndt* 586

CHEMISTRY.—ION ASSOCIATION IN AQUEOUS SOLUTIONS OF 1-1 ELECTROLYTES AT 25° . . . . . *By Mario Azzarri and Charles A. Kraus* 590

CHEMISTRY.—THE OPTICAL ROTATORY DISPERSION OF SIMPLE POLYPEPTIDES. I . . . . . *By William Moffitt and Jen Tsi Yang* 596

ENGINEERING.—ISOTROPIC MATERIALS WHOSE DEFORMATION AND DISTORTION ENERGIES ARE EXPRESSIBLE BY SCALAR INVARIANTS . . . . . *By T. Y. Thomas* 603

GENETICS.—CYTOLOGICAL DEMONSTRATION OF THE HELICAL STRUCTURE OF GIANT CHROMOSOMES . . . . . *By William Hovanitz* 609

GENETICS.—HETEROCARYOSIS AND PROTOPLASMIC INCOMPATIBILITY IN *Neurospora Crassa* . . . . . *By Laura Garnjobst and James F. Wilson* 613

GENETICS.—THE ORIENTATION AND EXTENT OF GENE TRANSFER IN *Escherichia Coli* . . . . . *By P. D. Skaar and Alan Garen* 619

MATHEMATICS.—A PROPERTY OF SPECIAL JORDAN ALGEBRAS . . . . *By A. A. Albert* 624

MATHEMATICS.—ON A CLASS OF FUNCTIONAL EQUATIONS OF MODULAR TYPE . . . . . *By Richard Bellman* 626

MATHEMATICS.—ON THE PRINCIPLE OF INVARIANT IMBEDDING AND PROPAGATION THROUGH INHOMOGENEOUS MEDIA . . . . . *By Richard Bellman and Robert Kalaba* 629

MATHEMATICS.—ON MULTIPLICATIVE IDEMPOTENTS OF A POTENT SEMIRING . . . . . *By Samuel Bourne* 632

MATHEMATICS.—GENERALIZED RESOLUTIONS OF THE IDENTITY FOR CLOSED SYMMETRIC ORDINARY DIFFERENTIAL OPERATORS . . . . . *By Earl A. Coddington* 638

MATHEMATICS.—CAUCHY'S PROBLEM FOR LINEAR DIFFERENTIAL EQUATIONS WITH CONSTANT COEFFICIENTS . . . . . *By Leon Ehrenpreis* 642

MATHEMATICS.—A UNIFIED THEORY OF GENERAL RELATIVITY OF GRAVITATION AND ELECTROMAGNETISM. II . . . . . *By Luther P. Eisenhart* 646

MATHEMATICS.—SINGULAR INTEGRAL EQUATIONS FOR DIFFERENTIAL FORMS ON RIEMANNIAN MANIFOLDS . . . . . *By J. J. Kohn* 650

MATHEMATICS.—MEASURES OF THE VALUE OF INFORMATION . . . *By John McCarthy* 654

MATHEMATICS.—DIOPHANTINE EQUATIONS IN CERTAIN RINGS . . . *By H. S. Vandiver* 656

PHYSICS.—A FUNDAMENTAL THEORY OF SUPERCONDUCTIVITY . . . *By Kenneth S. Pitzer* 665

PHYSICS.—CONVERGENT APPROACH TO THE THEORY OF ELEMENTARY PARTICLE INTERACTIONS . . . . . *By I. E. Segal* 670

PHYSIOLOGY.—TEMPERATURE INDEPENDENCE IN A UNICELLULAR "CLOCK" . . . . . *By Victor G. Bruce and Colin S. Pittendrigh* 676

PHYSIOLOGY.—INFLUENCE OF PROTEIN AND CALCIUM ADDITIONS TO ALREADY ADEQUATE DIET . . . . . *By Henry C. Sherman, Constance S. Pearson, Mary E. R. Bal, Andrea McCarthy, and Caroline Sherman Lanford* 682

PHYSIOLOGY.—MEASUREMENT OF IMPOSED VOLTAGE GRADIENT ADEQUATE TO MODULATE NEURONAL FIRING . . . . . *By C. A. Terzuolo and T. H. Bullock* 687

ZOÖLOGY.—IN VITRO AND IN VIVO EFFECTS OF A NERVE GROWTH-STIMULATING AGENT ISOLATED FROM SNAKE VENOM . . . . *By Rita Levi-Montalcini and Stanley Cohen* 695

ZOÖLOGY.—ON TIME LAGS IN EQUATIONS OF GROWTH . . . . . *By P. J. Wangersky and W. J. Cunningham* 699

ERRATUM.—COURSE OF PRODUCTION OF AN ISOANTISERUM EFFECTING TUMOR HOMOGRAFT SURVIVAL IN MICE . . . . . *By Nathan Kaliss* 702